

# Stochastic maximum principle for optimal control of a class of nonlinear SPDEs with dissipative drift \*

Marco Fuhrman <sup>†</sup>

Dipartimento di Matematica, Politecnico di Milano  
via Bonardi 9, 20133 Milano, Italia

Carlo Orrieri <sup>‡</sup>

Dipartimento di Matematica, Università di Pavia  
via Ferrata 1, 27100 Pavia, Italia

## Abstract

We prove a version of the stochastic maximum principle, in the sense of Pontryagin, for the finite horizon optimal control of a stochastic partial differential equation driven by an infinite dimensional additive noise. In particular we treat the case in which the non-linear term is of Nemytskii type, dissipative and with polynomial growth. The performance functional to be optimized is fairly general and may depend on point evaluation of the controlled equation. The results can be applied to a large class of non-linear parabolic equations such as reaction-diffusion equations.

## 1 Introduction

In the framework of classical optimal control theory, the maximum principle in the sense of Pontryagin is generally understood as a necessary condition for optimality of a control and its associated trajectory. Under some additional assumptions, the conditions may also become sufficient to ensure optimality. In the context of stochastic systems, a very general formulation of the maximum principle was obtained by Peng in [23], in the case of a controlled finite dimensional equation driven by the Brownian motion. This work was later extended in several directions by many authors. Here we pay particular attention to the case of infinite dimensional controlled systems, in particular controlled stochastic partial differential equations (SPDEs), on a finite time horizon.

The first result in infinite dimensions is due to Bensoussan [4] in the case of a convex set of control actions and a possibly infinite-dimensional Brownian noise. Later in [17] the case of diffusion coefficient unaffected by the control parameter was treated. Several versions of the stochastic maximum principle for general evolution equations were proved in [24], other versions can be found in [13], [12] and [19], under various conditions. The results in [14], [15] can be directly applied to a large class of concrete controlled SPDEs of parabolic type,

---

\*Financial support from the grant MIUR-PRIN 2010-11 “Evolution differential problems: deterministic and stochastic approaches and their interactions” is gratefully acknowledged. The second author have been supported by the Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM).

<sup>†</sup>Email: marco.fuhrman@polimi.it

<sup>‡</sup>Email: carlo.orrieri01@ateneopv.it

but only in the case of a finite-dimensional Brownian noise. It is worth emphasizing that the general case (in which the control parameter enters the diffusion term, the set of control actions is general—in particular non convex—and the noise is infinite-dimensional) is still an open problem.

The aim of this work is to prove a stochastic maximum principle for a class of controlled semilinear SPDEs of reaction-diffusion type written formally as

$$\begin{cases} \frac{\partial X}{\partial t}(t, \xi) = \mathcal{A}X(t, \xi) + f(X(t, \xi), u(t)) + B \frac{\partial^2 w}{\partial \xi \partial t}(t, \xi) \\ X(0, \xi) = x_0(\xi), \\ X(t, \xi) = 0 \quad \text{on} \quad \partial \mathcal{O}, \end{cases} \quad (1)$$

where  $\xi \in \mathcal{O}$  a bounded subset of  $\mathbb{R}^d$ , and  $t \in [0, T]$  for some fixed  $T > 0$ .  $u(\cdot)$  is the control process, taking values in a general space of control actions  $U$ ,  $\mathcal{A}$  is an operator of elliptic type,  $x_0$  a given initial condition,  $\partial^2 w / \partial \xi \partial t$  is a symbolic notation for a space-time white noise,  $B$  specifies the covariance operator of the noise, which enters the equation in an additive way, and finally  $f : \mathbb{R} \times U \rightarrow \mathbb{R}$  is a given function. Thus, the non-linear part of the drift is the so-called Nemytskii operator (or superposition operator) associated to  $f$ .

In the context of the general theory of SPDEs, the well-posedness of this type of equation (for a given control  $u(\cdot)$ ) is known under very general conditions on the function  $f$ , see for instance [8] for a systematic exposition. Particular attention is paid to the case when  $f$  is decreasing, which corresponds to a dissipativity assumption on the non-linear part of the drift, but it is otherwise very general, in particular it does not satisfy any Lipschitz condition. Most of the existing results were proved with the primary aim of studying the asymptotic behavior of an uncontrolled SPDE. In [6] dissipativity assumptions are required on the drift in connection with optimal control of an SPDE, with the aim to prove existence of relaxed optimal controls. The analysis is often carried out by re-writing the SPDE as an abstract evolution equation whose trajectories lie in the space  $H := L^2(\mathcal{O})$  or, under appropriate conditions, in the space  $E := C(\bar{\mathcal{O}})$ .

On the contrary all of the previously mentioned references on the stochastic maximum principle require the function  $f$  to be at least Lipschitz continuous. It is our purpose to prove a version of the stochastic maximum principle by imposing to the nonlinear term only dissipativity and polynomial growth conditions, along with some smoothness property. Even for finite dimensional controlled stochastic equations, this seems to be an issue that was considered only very recently: see [21]. On the other hand, we limit ourselves to the case where the noise is additive and uncontrolled.

In order to achieve greater generality we formulate our results for an abstract stochastic controlled evolution equation with additive noise in the Banach space  $E$ , of the form

$$\begin{cases} dX(t) = [AX(t) + F(X(t), u(t))] dt + BdW(t), \\ X(0) = x_0. \end{cases}$$

We assume that  $A$  is a dissipative sectorial operator in  $E$ , and thus the the generator of an analytic contraction semigroup of linear bounded operators  $S(t)$ ,  $t \geq 0$ . We assume that  $S$  extends to a semigroup on  $H$  satisfying in particular the regularizing property  $S(t)(H) \subset E$  for  $t > 0$ , as well as other assumptions. The non-linear term  $F : E \rightarrow E$  is of Nemytskii type and it is dissipative in  $E$ . The noise  $W$  is a cylindrical Wiener process in another separable Hilbert space  $K$  and  $B : K \rightarrow H$  is a linear bounded operator. Under suitable assumptions on  $B$  we can guarantee that the stochastic convolution  $\int_0^t S(t-s)BdW(s)$ ,  $t \in [0, T]$ , admits an  $E$ -continuous version and therefore by known results (see e.g. [10]) that the state equation admits a unique mild solution in  $E$  for any admissible control. We take special care in verifying that our abstract assumptions can be effectively checked in concrete cases like equation (1).

In our context the stochastic maximum principle is the statement that to any optimal pair  $(u, X)$  we can associate a pair of adjoint processes  $(p, q)$  in such a way that the Hamiltonian function, constructed by means of  $u, X$  and  $p$ , satisfies a maximum (or minimum) condition, see inequality (10) below. Our main results are two versions of the stochastic maximum principle, with slightly different assumptions. In the first one the adjoint processes  $(p, q)$  are defined and uniquely characterized using a duality argument. In the second one it is shown that  $(p, q)$  can also be characterized as the unique mild solution to a backward SPDE (BSPDE). While it is natural to expect that the first adjoint process  $p$  takes values in the dual space  $E'$  of  $E$ , it seems difficult to prove directly well-posedness of a BSPDE as an equation in  $E'$ , due to the lack of an efficient stochastic calculus in this space. To avoid this problem we formulate the BSPDE as an equation in a bigger Hilbert space where we can use standard stochastic calculus and then we prove that the solution is indeed more regular. In fact, under suitable assumptions we prove that the first adjoint process  $p$  actually takes values in  $H'$ , with a possible blow up when  $t \rightarrow T$ . In the proof we make frequent use of duality arguments in order to obtain a priori estimates.

We finally mention that developing the theory in the space  $E$  of continuous functions allows to consider fairly general cost functionals. For example we are able to formulate a control problem in which we optimize the evolution of the state evaluated at fixed points of the space, see (26).

The plan of the paper is as follows. After some preliminary results on the well-posedness of the state equation and the formulation of the optimal control problem in Section 2, we state our main results, Theorems 3.1 and 3.2, in Section 3. Section 4 is devoted to some examples, where we show that our general results can indeed be applied to various concrete controlled SPDEs. In Section 5 we analyze the spike variation technique and in Section 6 we prove Theorem 3.1. Finally, in Section 7 we study the BSPDE for the adjoint processes, which immediately leads to the proof of Theorem 3.2.

## 2 Notations, assumptions and preliminaries

Let  $\mathcal{O} \subset \mathbb{R}^d$  be an open bounded subset of  $\mathbb{R}^d$  with boundary  $\partial\mathcal{O}$  of class  $C^2$ . We denote by  $H$  the Hilbert space  $L^2(\mathcal{O}, \mathbb{R})$  with inner product  $\langle \cdot, \cdot \rangle_H$  and by  $E$  the Banach space  $C(\bar{\mathcal{O}}, \mathbb{R})$  endowed with the supremum norm  $|\cdot|_E$ . Moreover we denote by  ${}_{E'}\langle \cdot, \cdot \rangle_E$  (or simply  $\langle \cdot, \cdot \rangle_E$ ) the duality pairing in  $E' \times E$ , where  $E'$  is the topological dual of  $E$ , and by  ${}_{H'}\langle \cdot, \cdot \rangle_H$  the duality between  $H$  and  $H'$ . Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , by a cylindrical Wiener process we mean a family of linear mappings from  $K$  to  $L^2(\Omega)$

$$K \ni h \mapsto W_t^h \in L^2(\Omega)$$

such that the two following conditions hold

1.  $(W_t^h)_{t \geq 0}$  is a real continuous Wiener process, for every  $h \in K$ ;
2.  $\mathbb{E}(W_t^h \cdot W_t^k) = \langle h, k \rangle_K$ , for every  $h, k \in K$ .

**Remark 2.1.** We can also think the cylindrical Wiener process as

$$W(t) = \sum_{k=1}^{\infty} e_k W_k(t)$$

where  $e_k$  is a complete orthonormal system of  $K$  and  $W_k(t)$  are mutually independent real Brownian motions defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . It is worth noting that the series above does not converge in  $K$ , but in any Hilbert space  $K_1 \supset K$  such that the embedding is Hilbert-Schmidt (see [10] for a detailed exposure).

We use the natural filtration  $(\mathcal{F}_t)_{t \geq 0}$  associated to  $W$ , augmented in the usual way with the family of  $\mathbb{P}$ -null sets of  $\mathcal{F}$ , and we denote by  $\mathcal{P}$  the progressive  $\sigma$ -algebra on  $\Omega \times [0, T]$ , for some  $T > 0$ . If  $B$  is any Banach space, for any  $p \geq 1$ ,  $\lambda \in \mathbb{R}$  and  $T > 0$  we define

- $L_{\mathcal{F}_T}^p(\Omega; B)$ , the set of all  $\mathcal{F}_T$ -measurable random variables  $Y$  with values in  $B$  such that

$$\|Y\|_{L_{\mathcal{F}_T}^p(\Omega; B)} = (\mathbb{E}|Y|_B^p)^{1/p} < \infty;$$

- $L_{\mathcal{F}}^2(\Omega \times [0, T]; B)$ , the set of all  $(\mathcal{F}_t)$ -progressive processes with values in  $B$  such that

$$\|X\|_{L_{\mathcal{F}}^2(\Omega \times [0, T]; B)} = \left( \mathbb{E} \int_0^T |X(t)|_B^2 dt \right)^{1/2} < \infty;$$

- $L_{\mathcal{F}}^p(\Omega; C([0, T]; B))$  the set of all  $(\mathcal{F}_t)$ -adapted continuous processes with values in  $B$  such that

$$\|X\|_{L_{\mathcal{F}}^p(\Omega; C([0, T]; B))} = \left( \mathbb{E} \left( \sup_{t \in [0, T]} |X|_B \right)^p \right)^{1/p}$$

- $L_{\mathcal{F}}^p(\Omega; L^2([0, T], \lambda; B))$ , the set of all  $(\mathcal{F}_t)$ -progressive processes with values in  $B$  such that

$$\|X\|_{L_{\mathcal{F}}^p(\Omega; L^2([0, T], \lambda; B))} = \left( \mathbb{E} \left( \int_0^T |X(t)|_B^2 (T-t)^\lambda dt \right)^{p/2} \right)^{1/p} < \infty.$$

**Remark 2.2.** If  $B$  is a Hilbert space then also the spaces  $L_{\mathcal{F}}^2([0, T]; B)$  and  $L_{\mathcal{F}}^2([0, T], \lambda; B)$  are Hilbert. Moreover, it is easy to see that

$$[L_{\mathcal{F}}^2([0, T], \lambda; B)]' = L_{\mathcal{F}}^2([0, T], -\lambda; B)$$

where the duality is given by  $\langle x, y \rangle = \mathbb{E} \int_0^T \langle x(t), y(t) \rangle dt$ .

If  $K, H$  are Banach spaces we denote  $\mathcal{L}(K, H)$  the space of linear bounded operators  $T$  from  $K$  to  $H$ , endowed with the usual operator norm  $\|T\|_{\mathcal{L}(K, H)}$ . We set  $\mathcal{L}(K) := \mathcal{L}(K, K)$ . When  $K, H$  are Hilbert spaces we write  $\mathcal{L}_2(K, H)$  for the space of Hilbert-Schmidt operators from  $K$  to  $H$ , that is, linear operators  $T \in \mathcal{L}(K, H)$  such that

$$\|T\|_{\mathcal{L}_2(K, H)}^2 = \sum_{k \in \mathbb{N}} |Te_k|^2 < \infty$$

where  $(e_k)_{k \in \mathbb{N}}$  is any orthonormal basis of  $K$ . We set  $\mathcal{L}_2(K) := \mathcal{L}_2(K, K)$ .

The space of control actions is a general separable metric space  $U$  endowed with its Borel  $\sigma$ -algebra  $\mathcal{B}(U)$ . A control process is a progressive process  $(u_t)_{t \in [0, T]}$  with values in  $U$ . We denote with  $\mathcal{U}$  the space of admissible controls.

## 2.1 First assumptions on the controlled state equation

The aim of this work is to study a controlled stochastic partial differential equation of the form

$$\begin{cases} dX(t) = [AX(t) + F(X(t), u(t))] dt + BdW(t) \\ X(0) = x, \end{cases} \quad (2)$$

and try to give some necessary conditions for the existence of a control process which minimize a cost functional of the following type

$$J(u) = \mathbb{E} \int_0^T L(t, X(t), u(t)) dt + \mathbb{E} G(X(T)). \quad (3)$$

A control process  $\bar{u}$  for which (3) attains its minimum is called optimal, i.e.

$$J(\bar{u}) = \inf_{u(\cdot) \in \mathcal{U}} J(u(\cdot)).$$

$\bar{u}$ , together with its corresponding trajectory  $\bar{X}$ , will be called an optimal pair  $(\bar{u}, \bar{X})$ .

Let us give some assumptions on the equation we are considering.

**Hypothesis 2.1.** 1.  $A : D(A) \subset E \rightarrow E$  is a sectorial operator in  $E$ , which generates a contraction semigroup  $S(t)$ ,  $t \geq 0$ . Moreover,  $S(\cdot)$  extends to a strongly continuous semigroup in  $H$ , still denoted with the same symbol. We assume that  $S(t)(H) \subset E$  for  $t > 0$ , and for some constants  $C \geq 0$  and  $\lambda \in [0, 1)$  it holds that

$$\|S(t)\|_{\mathcal{L}(H,E)} \leq \frac{C}{t^\lambda}, \quad t \in (0, T]. \quad (4)$$

2.  $B \in \mathcal{L}(K, H)$ ,  $S(t)B \in \mathcal{L}_2(K, H)$  for almost every  $t \in [0, T]$ ,

$$\int_0^T \|S(t)B\|_{\mathcal{L}_2(K,H)}^2 dt < \infty \quad (5)$$

and the stochastic convolution

$$W_A(t) := \int_0^t S(t-s)BdW(s), \quad t \in [0, T],$$

admits a modification with trajectories being continuous functions with values in  $E$ .

3.  $x \in E$ .

**Remark 2.3.** 1. Since  $A$  is sectorial, it generates an analytic semigroup  $S(\cdot)$  of bounded linear operators in  $E$ , non necessarily strongly continuous, and  $D(A)$  may not be dense in  $E$  (see e.g. [20]). We additionally require that the semigroup is contractive, that is it satisfies  $\|S(t)\|_{\mathcal{L}(E)} \leq 1$  for every  $t \geq 0$ , and that it has an extension as required above.

2. The property (4) is a form of ultracontractivity, with the additional quantitative requirement that  $\lambda < 1$ . We will use this property in Proposition 6.1.

3. Under (5), for every  $t \in [0, T]$ , the stochastic integral  $W_A(t)$  is well defined as an equivalence class of random variables with values in  $H$ . We require that there exists an  $E$ -valued continuous stochastic process that, considered as a process in  $H \supset E$ , is a modification of  $W_A$ .

4. All the results that follow admit easy generalizations to the case when the drift  $F$  is time dependent or even stochastic. Similarly, the initial condition  $x \in E$  might be replaced by a random variable. We keep the previous setting to simplify the notation.

In order to state Theorems 3.1 and 3.2 below, we need to require that the semigroup is indeed more regular. To do it, we formulate two different hypotheses

**Hypothesis 2.2.** *There exists a Hilbert space  $V$ , continuously and densely embedded in  $E$ , and a constant  $\beta$  such that*

$$S(t)(V) \subset V, \quad \|S(t)\|_{\mathcal{L}(V)} \leq \beta, \quad t \in (0, T].$$

**Hypothesis 2.3.** *There exists a Hilbert space  $V$ , continuously and densely embedded in  $E$ , and a constant  $\beta$  such that*

$$S(t)(E) \subset V, \quad \|S(t)\|_{\mathcal{L}(V)} \leq \beta, \quad t \in (0, T].$$

The second assumption is clearly more stringent. Note however that in both cases the bound concerns the norm in  $\mathcal{L}(V)$ . It will be clear later on (see Proposition 6.1 and Section 7) which are the motivations for the introduction of the new space  $V$ .

Regarding the cost functional, we need the following

- Hypothesis 2.4.** 1. *The functionals  $L : \Omega \times [0, T] \times E \times U \rightarrow \mathbb{R}$  and  $G : \Omega \times E \rightarrow \mathbb{R}$  are measurable with respect to  $\mathcal{P} \otimes \mathcal{B}(E) \otimes \mathcal{B}(U)$  and  $\mathcal{B}(\mathbb{R})$  (respectively  $\mathcal{F}_T \otimes \mathcal{B}(E)$  and  $\mathcal{B}(\mathbb{R})$ ).*
2. *For each  $(t, u) \in [0, T] \times U$ , the operators  $L$  and  $G$  are Fréchet differentiable on  $E$ . Moreover there exist  $K \geq 0$ ,  $k \geq 0$  such that,  $\mathbb{P}$ -a.s.*

$$|L(t, x, u)| + |D_x L(t, x, u)|_E \leq K(1 + |x|_E^k), \quad t \in [0, T], x \in E, u \in U,$$

$$|D_x G(x)|_E \leq K(1 + |x|_E^k), \quad x \in E.$$

## 2.2 Assumptions on the drift as a Nemytskii operator

Here we describe the non-linear term  $F$  in the state equation. We suppose that a function  $f : \mathbb{R} \times U \rightarrow \mathbb{R}$  is given and that  $F : E \times U \rightarrow E$  is defined by

$$F(x, u)(\xi) = f(x(\xi), u), \quad \xi \in \mathcal{O}$$

for any continuous map  $x : \bar{\mathcal{O}} \rightarrow \mathbb{R}$  and  $u \in U$ . Thus,  $F$  is the so-called *Nemytskii operator* associated to the real function  $f$ , on which we make the following assumptions.

- Hypothesis 2.5.** 1.  *$f : \mathbb{R} \times U \rightarrow \mathbb{R}$  is measurable. For every  $\sigma \in \mathbb{R}$  the map  $u \mapsto f(\sigma, u)$  is continuous on  $U$ , and for every  $u \in U$  the map  $\sigma \mapsto f(\sigma, u)$  is continuously differentiable on  $\mathbb{R}$ , with derivative denoted by  $f'(\sigma, u)$ . Moreover there exist  $C \geq 0$ ,  $k \geq 0$  such that*

$$|f(\sigma, u)| + |f'(\sigma, u)| \leq C(1 + |\sigma|^k), \quad \sigma \in \mathbb{R}, u \in U. \quad (6)$$

2. *There exists  $\beta \in \mathbb{R}$  such that*

$$f'(\sigma, u) \leq \beta, \quad \sigma \in \mathbb{R}, u \in U.$$

If the map  $f$  satisfies the above assumption, the Nemytskii operator  $F$  cannot be defined as a map from  $H \times U$  into  $H$ . Even if it could be defined, under more restrictive growth assumptions on  $f$ , it would fail to be differentiable in general: indeed, it is known that a Nemytskii operator  $T : H \rightarrow H$  is Fréchet differentiable if and only if  $T$  is linear (see [2], page 20). However, as a map from  $E \times U$  to  $E$ ,  $F$  is well defined and has some regularity properties, as shown in the following

**Lemma 2.6.** *The functional  $F : E \times U \rightarrow E$  is Fréchet differentiable on  $E$ . Moreover the Fréchet differential  $D_x F$  acts on every  $h \in E$  as a multiplication operator, namely*

$$[D_x F(x, u) \cdot h](\xi) = f'(x(\xi), u) \cdot h(\xi), \quad \xi \in \mathcal{O}. \quad (7)$$

*Proof.* Fix  $x, h \in E$ ,  $u \in U$ . For  $\xi \in \mathcal{O}$ ,  $s \neq 0$  we have

$$f(x(\xi) + sh(\xi), u) - f(x(\xi), u) = \int_0^1 f'(x(\xi) + \theta sh(\xi), u) sh(\xi) d\theta. \quad (8)$$

Fix  $\varepsilon > 0$ , let  $\delta > 0$  be such that  $|f'(x(\xi) + y, u) - f'(x(\xi), u)| < \varepsilon$ , for  $|y| \leq \delta$ . Such a  $\delta$  exists since  $f'(\cdot, u)$  is continuous on a compact set containing the image of  $x(\cdot)$ . Now choose  $s$  small enough such that  $|\theta sh(\xi)| \leq \delta$ . Taking the limit we obtain

$$\begin{aligned} & \sup_{\xi \in \mathcal{O}} \frac{1}{s} |f(x(\xi) + sh(\xi), u) - f(x(\xi), u) - f'(x(\xi), u) sh(\xi)| \\ &= \sup_{\xi \in \mathcal{O}} \left| \int_0^1 [f'(x(\xi) + \theta sh(\xi), u) - f'(x(\xi), u)] d\theta h(\xi) \right| \\ &\leq \varepsilon \sup_{\xi \in \mathcal{O}} |h(\xi)| \end{aligned}$$

which proves (7).  $\square$

To proceed further we need to recall the following

**Definition 2.1.** A map  $g : D(g) \subset E \rightarrow E$  is called *dissipative* if for all  $x, y \in D(g)$  and  $\alpha \geq 0$  it holds

$$|x - y|_E \leq |x - y - \alpha(g(x) - g(y))|_E.$$

Equivalently, if there exists  $z^* \in \partial|x - y|_E$  such that

$${}_E \langle g(x) - g(y), z^* \rangle_{E'} \leq 0 \quad \forall x, y \in D(g).$$

Here we denote by  $\partial|\cdot|_E$  the sub-differential of the norm in  $E$ .

In the special case in which  $E$  is a Hilbert space, this condition coincides with the monotonicity assumption  $\langle g(x) - g(y), x - y \rangle_E \leq 0$ , for all  $x, y \in D(g)$ .

We refer e.g. to [7] for basic properties of dissipative mappings. We also take from [7], page 180, the following two results on the properties of the operator  $F$  and its Yosida approximations.

**Lemma 2.7.** Under Hypothesis 2.5, if  $F$  is the Nemytskii operator associated to  $f$  then there exists a constant  $c \in \mathbb{R}$  such that the following are true:

- (i)  $F(\cdot, u) - cI$  is dissipative in  $E$ , for every  $u \in U$ ;
- (ii) If  $x \in E$  then  $\sup_{u \in U} \|D_x F(x, u)\|_{\mathcal{L}(E)} \leq c(1 + |x|_E^k)$ ;
- (iii) If  $x, h \in E$  then there exists  $\delta_h \in \partial|h|_E$  such that  $\sup_{u \in U} \langle D_x F(x, u)h, \delta_h \rangle_E \leq c|h|_E$ ;
- (iv) For  $x, h \in E$  it holds  $\sup_{u \in U} \langle D_x F(x, u)h, h \rangle_H \leq c|h|_H^2$ .

**Lemma 2.8.** For any  $\alpha > 0$ ,  $\sigma \in \mathbb{R}$  and  $u \in U$ , consider the resolvent map  $J_\alpha(\sigma, u) = (I - \alpha f)^{-1}(\sigma, u)$  and define  $f_\alpha(\sigma, u) = f(J_\alpha(\sigma, u), u) - cJ_\alpha(\sigma, u)$ . If we denote  $F_\alpha(x, u)(\xi) = f_\alpha(x(\xi), u)$ , then  $F_\alpha$  is dissipative and Lipschitz-continuous both in  $H$  and in  $E$ . Moreover

$$[DF_\alpha(x, u)y](\xi) = f'_\alpha(x(\xi), u) \cdot y(\xi), \quad \xi \in \mathcal{O}, u \in U.$$

and for any  $R > 0$  we have

$$\lim_{\alpha \rightarrow 0} \sup_{|x|_E \leq R} |F_\alpha(x, u) - F(x, u)|_E = 0, \quad u \in U.$$

**Remark 2.4.** Let us notice that the operators  $F_\alpha$  defined above coincide with the usual Yosida approximations of  $F$ . We also note that from (6) it follows that

$$|f_\alpha(\sigma, u)| + |f'_\alpha(\sigma, u)| \leq C(1 + |\sigma|^k), \quad \sigma \in \mathbb{R}, u \in U. \quad (9)$$

for some  $C > 0$  independent of  $\alpha$ .

### 2.3 The state equation

Now we are in position to study the abstract form of the state equation (2). While (2) is merely a formal writing, the precise formulation of the state equation is a so-called *mild formulation*: an  $E$ -valued continuous adapted process  $X(t)$  is a *mild solution* of the SPDE above if,  $\mathbb{P}$ -a.s.,

$$X(t) = S(t)x + \int_0^t S(t-s)F(X(s), u(s))ds + W_A(t), \quad t \in [0, T],$$

where  $W_A$  is the  $E$ -continuous modification of  $\int_0^t S(t-s)BdW(s)$ ,  $t \in [0, T]$ . Under the previous assumptions we can state the following theorem (compare [10]).

**Theorem 2.9.** *Assume that hypotheses 2.1 and 2.5 hold true, then equation (2) admits a unique mild solution  $X$ . Moreover  $X \in L^p_{\mathcal{F}}(\Omega; C([0, T]; E))$  for every  $p \geq 1$ .*

## 3 Statement of the main results

Let  $V'$  be the dual of the Hilbert space  $V$  introduced in Hypothesis 2.2. Let us identify  $H$  with  $H'$  by the Riesz isometry. Then we obtain the following continuous dense inclusions

$$V \subset E \subset H \subset E' \subset V'.$$

Now we can state the two building blocks of our main result.

**Theorem 3.1.** *Suppose Hypotheses 2.1, 2.2, 2.4 and 2.5 hold, let  $r' \in (1, 2)$  be fixed arbitrarily and let  $\lambda$  be as in Hypothesis 2.1. Suppose that  $(u, X)$  is an optimal pair. Then there exists a progressive process  $p$  with values in  $H'$  satisfying*

$$\mathbb{E} \left( \int_0^T |p(t)|_{H'}^2 (T-t)^\lambda dt \right)^{r'} < \infty$$

and for which the following inequality holds  $\mathbb{P}$ -a.s. for almost every  $t \in [0, T]$ :

$$\mathcal{H}(t, v, X(t), p(t)) - \mathcal{H}(t, u(t), X(t), p(t)) \geq 0, \quad \text{for every } v \in U, \quad (10)$$

where  $\mathcal{H}(t, u, x, p) := L(t, x, u) + {}_{H'}\langle p, F(x, u) \rangle_H$  is the Hamiltonian function of the system.

We note that in Theorem 3.1 no characterization of  $p$  is given. It will be clear from the proof that  $p$  is the first component of a pair  $(p, q)$  which is uniquely determined by a duality argument. In Theorem 3.2 we give a more precise characterization of the pair  $(p, q)$  as the solution of an adjoint equation, which is backward in time.

**Theorem 3.2.** *Let  $(u, X)$  be an optimal pair. If Hypotheses 2.1, 2.3, 2.4 and 2.5 hold and  $r' \in (1, 2)$  then there exists a pair of progressive processes  $(p, q)$  with values in  $H' \times L_2(K, V')$  for which inequality (10) holds. Moreover*

$$\mathbb{E} \left( \int_0^T |p(t)|_{H'}^2 (T-t)^\lambda dt \right)^{r'} + \mathbb{E} \left( \int_0^T |q(t)|_{V'}^2 dt \right)^{r'} < \infty.$$

and the pair  $(p, q)$  is the unique mild solution of the Backward SPDE (53) below.

**Remark 3.1.** It is worth noting that Hypothesis 2.3 is used only in the proof of uniqueness of the Backward SPDE. For the existence part it is sufficient that the semigroup preserves  $V$ , i.e. Hypothesis 2.2.



## 4 Examples

The aim of this section is to present some concrete examples of stochastic control problems for SPDEs which can be treated using our results and where the general assumptions stated above can be effectively checked. We begin by giving some general sufficient conditions for the space-time continuity of the stochastic convolution, that we have assumed in Hypothesis 2.1-2. Then we focus on the case in which the sectorial operator  $A$  is a realization of the Laplace operator with Dirichlet boundary conditions: we give examples of covariance operators  $B$  for which Hypothesis 2.1-2 can be readily verified and we show how the abstract Hilbert space  $V$  can be chosen. Last, we briefly discuss the form of the cost functionals in which we are more interested in.

### 4.1 A general condition for the space-time continuity of $W_A(t)$

In the literature, most of the sufficient condition for space-time continuity of the stochastic convolution can only be applied to one-dimensional domains  $\mathcal{O} \subset \mathbb{R}$  and to few simple domains in higher dimensions. Here we prove some statements concerning more general domains in  $\mathbb{R}^n$ ,  $n > 1$ , with boundary regular enough, mainly relying on the results from [16]. Although several arguments are usual, we could not find any reference that includes the results we are going to prove.

The first step in this direction is the study of the Ornstein-Uhlenbeck process associated with the stochastic equation

$$dz_t = Az_t dt + BdW_t. \quad (11)$$

The existence of mild solution  $W_A(t)$  in  $H$  to the above equation is easily obtained by imposing the following trace condition, for any  $t \geq 0$ ,

$$\int_0^t \|S(s)B\|_{L_2(H)} ds = \int_0^t \text{Tr} [S(s)BB^*S(s)^*] ds < \infty. \quad (12)$$

Under this assumption,  $W_A(t)$  is a mean-square continuous Gaussian process with values in  $H$  and it is given by

$$W_A(t) = \int_0^t S(t-s)BdW(s). \quad (13)$$

For our purposes, we need  $W_A$  to be more regular and we make the following

**Hypothesis 4.1.** (i) For any  $p \geq 2$ , the semigroup  $S(t)$  extends uniquely to a strongly continuous semigroup in  $L^p(\mathcal{O})$ .

(ii) For all  $\varepsilon \in (0, 1)$  and  $p \geq 2$ , there exist  $r \geq 2$  and  $K$  such that the following condition holds

$$|S(t)|_{W^{\varepsilon,p}(\mathcal{O})} \leq Kt^{-\frac{\varepsilon}{r}} |x|_{L^p(\mathcal{O})}, \quad \forall x \in L^p(\mathcal{O}),$$

where with  $W^{\varepsilon,p}$  we denoted the fractional Sobolev space.

(iii) There exists an orthonormal basis  $e_k$  of  $H$  which diagonalize both  $A$  and  $B$ . That is, there exist two sequences of real positive numbers  $\mu_k$  and real  $b_k$  such that  $\mu_k \nearrow +\infty$  as  $k \rightarrow +\infty$  and

$$Ae_k = -\mu_k e_k, \quad Be_k = b_k e_k, \quad k \in \mathbb{N};$$

(iv) Each  $e_k(\cdot)$  is bounded. Let  $M > 0$  and  $c_k$  be an increasing sequence, then it holds that

$$|e_k(\xi)| \leq M c_k, \quad c_k \in \mathbb{N}, \xi \in \mathcal{O};$$

(v) There exists  $\alpha \in (0, \frac{1}{2})$  such that

$$\sum_{k=1}^{+\infty} b_k^2 \mu_k^{2\alpha-1} c_k^2 < \infty$$

Now we are in position to state the following

**Theorem 4.2.** *Assume Hypotheses (i)-(v) hold. Then the stochastic convolution  $W_A : [0, +\infty) \times \mathcal{O} \rightarrow \mathbb{R}$  is continuous,  $\mathbb{P}$ -a.s. Moreover, if  $p \geq 2$  we have*

$$\mathbb{E} \sup_{t \in [0, T]} |W_A(t)|_E^p < \infty$$

Before proving the theorem we recall a useful analytic lemma (cfr. [9], page 23).

**Lemma 4.3.** *Assume Hypotheses (i)-(ii) hold. Let  $T > 0$ ,  $p \geq 2$  and  $f \in L^p([0, T] \times \mathcal{O})$ . If we set*

$$F(t) = \int_0^t S(t-\sigma)(t-\sigma)^{\alpha-1} f(\sigma) d\sigma, \quad t \in [0, T];$$

*then  $F \in C([0, T] \times \mathcal{O})$  and there exists a constant  $C_{T,p}$  such that the following estimate hold*

$$\sup_{t, \xi} |F(t, \xi)|^p \leq C_{T,p} |f|_{L^p([0, T] \times \mathcal{O})}^p.$$

*Proof of Theorem 4.2.* Using the factorization method (see e.g. [10] for a detailed exposition), we write  $W_A(t)$  in the following form

$$W_A(t) = \frac{\sin(\pi\alpha)}{\pi} \int_0^t S(t-\sigma)(t-\sigma)^{\alpha-1} Y(\sigma) d\sigma, \quad (14)$$

where

$$Y(\sigma) = \int_0^\sigma S(\sigma-s)(\sigma-s)^{-\alpha} B dW(s), \quad \sigma \geq 0. \quad (15)$$

If  $\xi \in \mathcal{O}$ , setting  $Y(\sigma)(\xi) =: Y(\sigma, \xi)$  we have

$$Y(\sigma, \xi) = \sum_{k=1}^{\infty} b_k \int_0^\sigma e^{-\mu_k(\sigma-s)} (\sigma-s)^{-\alpha} e_k(\xi) dW_k(s)$$

which is a gaussian random variable with zero mean and covariance given by

$$v_\alpha(\sigma, \xi) = \sum_{k=1}^{\infty} b_k^2 \int_0^\sigma e^{-2\mu_k s} s^{-2\alpha} |e_k(\xi)|^2 ds. \quad (16)$$

Thanks to Hypotheses (iv)-(v) we see that

$$v_\alpha(\sigma, \xi) \leq M \sum_{k=1}^{\infty} b_k^2 \int_0^\infty e^{-z} \frac{z^{-2\alpha}}{2\mu_k^{1-2\alpha}} c_k^2 dz \leq M \sum_{k=1}^{\infty} b_k^2 \mu_k^{2\alpha-1} c_k^2 < \infty. \quad (17)$$

Therefore, from the gaussianity of  $Y(\cdot, \cdot)$ , there exists  $C_p > 0$  such that

$$\mathbb{E} |Y(\sigma, \xi)|^p \leq c(\mathbb{E} |Y(\sigma, \xi)|^2)^{p/2} < C_p$$

and we have

$$\mathbb{E} \int_0^T \int_{\mathcal{O}} |Y(\sigma, \xi)|^p m(d\xi) dt \leq T C_p m(\mathcal{O}).$$

The conclusion follows from Lemma 4.3. □

## 4.2 The Laplace operator

Let us choose  $A$  to be the realization of the Laplace operator  $\Delta_\xi$  with Dirichlet boundary conditions. If  $A$  is a suitable uniformly elliptic second order differential operator of negative type the same argument holds true with some modification. The state equation reads as

$$\begin{cases} \frac{\partial X}{\partial t}(t, \xi) = \Delta_\xi X(t, \xi) + f(X(t, \xi), u(t)) + B \frac{\partial^2 w}{\partial \xi \partial t}(t, \xi) \\ X(0, \xi) = x_0(\xi), \quad \xi \in \mathcal{O}, \\ X(t, \xi) = 0, \quad t > 0, \xi \in \partial \mathcal{O}. \end{cases}$$

In this case, the regularizing effect of the heat semigroup can be expressed by

$$|S(t)x|_{s,2} \leq ce^{-\frac{s}{2}t}(t \wedge 1)^{-\frac{s}{2}}|x|_H, \quad (18)$$

where  $|\cdot|_{s,2}$  is the norm in the fractional Sobolev space  $W^{s,2}(\mathcal{O}, \mathbb{R})$ . The Sobolev embedding theorem assures that  $W^{s,2}(\mathcal{O}, \mathbb{R}) \hookrightarrow C(\bar{\mathcal{O}}, \mathbb{R})$  if  $s > d/2$ . Then we have that

$$|S(t)x|_\infty \leq ce^{-\frac{d}{4}t}(t \wedge 1)^{-\frac{d}{4}}|x|_H, \quad (19)$$

In order to satisfy Hypothesis 2.1 we have to choose  $d \leq 3$ , therefore in this framework we are able to study stochastic equation of reaction-diffusion type only on bounded domains of dimensions 1, 2 or 3. Thanks to the estimate (18), a good choice for the Hilbert space  $V$  is the fractional Sobolev space  $W^{s,2}(\mathcal{O}, \mathbb{R})$ , for which Hypotheses 2.2 and 2.3 are easily satisfied. Hence we end up with a triplet of the form

$$W^{s,2}(\mathcal{O}, \mathbb{R}) \subset C(\bar{\mathcal{O}}, \mathbb{R}) \subset L^2(\mathcal{O}, \mathbb{R}).$$

Concerning the stochastic convolution  $W_\Delta(t)$ , we want to present here an explicit computation which finalizes the Hypotheses 4.1 presented above. Assumptions (i)-(ii) are easily verified with the choice  $r = 2$ . The crucial hypothesis is the fourth one. Indeed the constant  $M$  in (iv) can depends on the domain  $\mathcal{O}$ . According to a remark in [16], the two basic domains one has to deal with are the square and the ball, which correspond to the best and worst case, respectively, as far as the growth of normalized eigenfunctions is concerned. Let us begin taking  $\mathcal{O} = [0, \pi]^d$ , the eigenfunctions of the Laplace operator are

$$e_k(\xi) = \left(\frac{2}{\pi}\right)^{\frac{d}{2}} \sin(k_1 \xi) \cdots \sin(k_d \xi), \quad \xi \in [0, \pi]^d, |k|^2 = k_1^2 + \dots + k_d^2.$$

hence they are uniformly bounded from above by a constant, i.e.  $|e_k(\xi)| \leq M$ , for all  $k \in \mathbb{N}$ . On the other hand, for a general domain with smooth enough boundary, Grieser in [16] produce an estimate of form

$$\sup_{x \in \mathcal{O}} |e_k(x)| \leq M \mu_k^{\frac{d-1}{4}}.$$

It is worth noting that the bound given above is optimal in the case in which the domain is a ball,  $\mathcal{O} = \{x \in \mathbb{R}^d : |x| \leq 1\}$ . In the following we will refer to it as to the “worst” case.

In many applications the diffusion operator is written as power of  $A$ , more precisely  $B = (-A)^{-\gamma}$ , for some  $\gamma \geq 0$ . We are looking for some condition on  $\gamma$  implying the continuity in space and time of the stochastic convolution. The general idea is the following: in order to have the space-time continuity it is sufficient to colour the noise a little bit more, in other words it is sufficient to increase enough the exponent  $\gamma$ . To do that, we recall a result proved by Agmon, see [1], and we follow [5] for the verification of the trace condition. Let  $N(\mu)$  be the number of eigenvalues of  $-A$  not exceeding  $\mu$ , then we have an asymptotic estimate of the following type

$$N(\mu) = c\mu^{d/2} + o(\mu^{d/2}), \quad \text{for some } c > 0. \quad (20)$$

The relation above is useful in defining the measure  $\nu$  on the real interval  $[\mu_1, \infty)$ . Here  $\mu_1$  is the first (positive) eigenvalue such that

$$\nu(D) = \#\{k : \mu_k \in D\},$$

for every Borel subset  $D$  of  $[\mu_1, \infty)$ . It is obvious that  $\nu([\mu_1, \mu]) = N(\mu)$ . If we rewrite Hypothesis (v) in this setting we get, on one hand,

$$\sum_{k=1}^{\infty} \mu_k^{2\alpha-1-2\gamma+(d-1)/2} < \infty, \quad (21)$$

for the “worst” case in which the domain  $\mathcal{O}$  is a ball. On the other hand, If  $\mathcal{O}$  is a hypercube the condition is the following

$$\sum_{k=1}^{\infty} \mu_k^{2\alpha-1-2\gamma} < \infty, \quad (22)$$

Let us consider first the worst case, we get

$$\begin{aligned} & \sum_{k=1}^{\infty} \mu_k^{2\alpha-1-2\gamma+(d-1)/2} \\ &= \int_{\mu_1}^{\infty} \mu^{2\alpha-1-2\gamma+(d-1)/2} \nu(d\mu) \\ &= - \int_{\mu_1}^{\infty} \int_{\mu}^{\infty} \frac{d}{ds} s^{2\alpha-1-2\gamma+(d-1)/2} ds \nu(d\mu) \\ &= \int_{\mu_1}^{\infty} \int_{\mu_1}^{\infty} C s^{2\alpha-2-2\gamma+(d-1)/2} 1_{s \geq \mu} ds \nu(d\mu) \\ &= C \int_{\mu_1}^{\infty} s^{2\alpha-2-2\gamma+(d-1)/2} N(s) ds \\ &\leq C \int_{\mu_1}^{\infty} s^{2\alpha-2-2\gamma+(d-1)/2+d/2} ds, \end{aligned} \quad (23)$$

Where we used (20) in the last inequality. Obviously, Hypothesis (v) is satisfied if and only if the last integral is finite. That is, in terms of  $\gamma$ , if and only if

$$\gamma > \frac{2d-3}{4} + \alpha. \quad (24)$$

Proceeding in the same way for  $\mathcal{O} = [0, \pi]^d$  we obtain the following condition

$$\gamma > \frac{d-2}{4} + \alpha. \quad (25)$$

As an example, let us consider for simplicity the two dimensional case. Here we have

- $\gamma > \alpha$  if  $\mathcal{O} = [0, \pi] \times [0, \pi]$ ,
- $\gamma > 1/4 + \alpha$  if  $\mathcal{O} = \{x \in \mathbb{R}^2 : |x| \leq 1\}$ ,

with  $\alpha \in (0, 1/2)$ . It is interesting to notice that in [5] the authors obtained the condition  $\gamma > d/4 - 1/2$  (if  $d = 2$ , this imply  $\gamma > 0$ ) for the verification of the trace condition. To obtain the space-time continuity the condition become more restrictive. In the two dimensional case, one has to choose an exponent  $\gamma$  for the diffusion term that is strictly greater than  $1/4$ .

### 4.3 Cost functional

One of the motivations for the study of optimal control problems in the space of continuous functions is the possibility to consider a very large class of cost functionals. Here we are particularly interested in costs of Nemytskii type of the following form

$$J(u) = \mathbb{E} \int_0^T \int_{\mathcal{O}} l(t, X(t, \xi), u(t)) \mu(d\xi) dt + \mathbb{E} \int_{\mathcal{O}} g(X(T, \xi)) \mu(d\xi), \quad (26)$$

where  $\mu$  is a given measure on  $\mathcal{O} \subset \mathbb{R}^d$  and  $l, h$  satisfy the following conditions

1. The functions

$$l(\omega, t, \sigma, u) : \Omega \times [0, T] \times \mathbb{R} \times U \rightarrow \mathbb{R} \quad \text{and} \quad g(\omega, \sigma) : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$$

are assumed to be measurable with respect to  $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(U)$  and  $\mathcal{B}(\mathbb{R})$  (respectively  $\mathcal{F}_T \otimes \mathcal{B}(\mathbb{R})$  and  $\mathcal{B}(\mathbb{R})$ ).

2. The maps  $\sigma \mapsto l(\omega, t, \sigma, u)$  and  $r \mapsto g(\omega, \sigma)$  are  $C^1(\mathbb{R})$  and there exists  $K > 0, k \geq 0$  such that,  $\mathbb{P}$ -a.s.

$$\sup_{u \in U} \sup_{t \in [0, T]} |D_\sigma l(t, \sigma, u)| \leq K(1 + |\sigma|^k),$$

$$|D_\sigma g(\sigma)| \leq K(1 + |\sigma|^k).$$

The two important cases we can deal with are

- $\mu_1 = m$  is the Lebesgue measure;
- $\mu_2 = \sum_{i=1}^n a_i \delta_{\xi_i}$  is a linear combination of Dirac measures at points  $\xi_i \in \mathcal{O}$ .

Observe that a cost of Nemytskii type with  $\mu = \mu_2$  reduces to a sum of pointwise evaluations which is well defined in the Banach space  $E$  of continuous functions but not in the space  $H$ . In our case, the development of a theory of stochastic optimal control problems in a Banach framework allows to control the evolution of a stochastic reaction-diffusion equation only in a finite number of points.

## 5 Spike variation method

In this section we state preliminary results needed for the proof of Theorems 3.1 and 3.2. We use the classical approach based on spike variations (in time). Throughout this section we suppose that Hypotheses 2.1, 2.4, 2.5 hold true.

Let us consider an optimal control  $u(t)$  and the corresponding optimal trajectory  $X(t)$ . To shorten somehow the notation, in this section we write  $u_t, X_t$  instead of  $u(t), X(t)$  and we use similar notation for other processes. Let  $E_\varepsilon \subset [0, T]$  be a set of measure  $\varepsilon$  of the form  $[t_0, t_0 + \varepsilon]$ , for some  $t_0 \in (0, T)$ , then we can introduce the spike variation process

$$u_t^\varepsilon = \begin{cases} u_t, & \text{if } t \in [0, T] \setminus E_\varepsilon \\ w_t, & \text{if } t \in E_\varepsilon, \end{cases} \quad (27)$$

for some admissible control process  $w$ . The perturbed trajectory is denoted by  $X^\varepsilon$ . We are going to construct a new process  $Y^\varepsilon$  which is the mild solution to the following equation, known as first variation equation:

$$\begin{cases} \frac{d}{dt} Y_t^\varepsilon = AY_t^\varepsilon + D_x F(X_t, u_t) Y_t^\varepsilon + \delta^\varepsilon F_t \\ Y_0^\varepsilon = 0, \end{cases} \quad (28)$$

where we used the notation  $\delta^\varepsilon F_t := F(X_t, u_t^\varepsilon) - F(X_t, u_t)$ . More precisely we say that  $Y^\varepsilon$  is a mild solution to the above equation if for all  $t > 0$ ,  $\mathbb{P}$ -a.s. we have

$$Y_t^\varepsilon = \int_0^t S(t-s)(D_x F(X_s, u_s)Y_s^\varepsilon + \delta^\varepsilon F_s)ds.$$

Existence and uniqueness of a mild solution of equation (28) with values in  $E$  is well known, see e.g. [7] for a detailed exposition. Let us recall that, according to Lemma 2.7, for any  $y \in E$

$$\langle DF(X_t, u_t)y, \delta_y \rangle_E \leq c|y|_E, \quad \mathbb{P} - \text{a.s.} \quad (29)$$

where  $\delta_y$  is an element of the subdifferential of the norm  $\partial|y|_E$  and  $c \in \mathbb{R}$ . We will use the following notation throughout the paper:

$$|DF(X_t, u_t)y|_E \leq C_F(\omega)|y|_E, \quad \mathbb{P} - \text{a.s.} \quad (30)$$

where  $C_F(\omega)$  denotes a suitable random variable having finite moments of any order. Now we want to write the difference of the cost functional  $J(u^\varepsilon) - J(u)$  as a function of  $Y^\varepsilon$ , up to a negligible reminder. In order to do that we need some estimate of the difference  $X - X^\varepsilon - Y^\varepsilon$ . More precisely we state the following

**Lemma 5.1.** *If we define  $\xi_t^\varepsilon := X_t^\varepsilon - X_t$ , and  $\eta_t^\varepsilon := \xi_t^\varepsilon - Y_t^\varepsilon$ , then the following estimates hold, for  $k \geq 1$ ,*

$$(i) \quad \mathbb{E} \sup_{t \in [0, T]} |\xi_t^\varepsilon|_E^{2k} = O(\varepsilon^{2k}),$$

$$(ii) \quad \mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon|_E^{2k} = O(\varepsilon^{2k}),$$

$$(iii) \quad \mathbb{E} \sup_{t \in [0, T]} |\eta_t^\varepsilon|_E^{2k} = o(\varepsilon^{2k}).$$

*Proof.* (i). Let us denote for brevity  $z_t := \int_0^t e^{(t-s)A} Q dW(s)$  and consider  $v_t := X_t - z_t$ ,  $v_t^\varepsilon := X_t^\varepsilon - z_t$ . Then

$$v_t = \int_0^t S(t-s)F(v_s + z_s, u_s)ds, \quad v_t^\varepsilon = \int_0^t S(t-s)F(v_s^\varepsilon + z_s, u_s^\varepsilon)ds.$$

Let  $v^{n, \varepsilon}, v^n$  solutions to the corresponding equations with  $S(\cdot)$  replaced by  $S_n(\cdot)$ . Therefore  $v^{\varepsilon, n}$  and  $v^n$  are solutions of the equations

$$\frac{d}{dt}v_t^n = A_n v_t^n + F(v_t^n + z_t, u_t), \quad \frac{d}{dt}v_t^{n, \varepsilon} = A_n v_t^{n, \varepsilon} + F(v_t^{n, \varepsilon} + z_t, u_t^\varepsilon) \quad (31)$$

where  $A_n = nA(n-A)^{-1}$  are the Yosida approximations of the operator  $A$ . Then, if  $\delta_t \in \partial|v_t^{n, \varepsilon} - v_t^n|_E$ , we have

$$\begin{aligned} \frac{d^-}{dt}|v_t^{n, \varepsilon} - v_t^n|_E &\leq \langle A_n(v_t^{n, \varepsilon} - v_t^n) + F(v_t^{n, \varepsilon} + z_t, u_t^\varepsilon) - F(v_t^n + z_t, u_t), \delta_t \rangle_E \\ &\leq \langle F(v_t^{n, \varepsilon} + z_t, u_t^\varepsilon) - F(v_t^n + z_t, u_t^\varepsilon), \delta_t \rangle_E + \langle F(v_t^n + z_t, u_t^\varepsilon) - F(v_t^n + z_t, u_t), \delta_t \rangle_E \\ &\leq c|v_t^{n, \varepsilon} - v_t^n|_E + |F(v_t^n + z_t, u_t^\varepsilon) - F(v_t^n + z_t, u_t)|_E, \end{aligned}$$

from the contraction property of the semigroup and the dissipativity of the operator  $F$ . Thanks to the Gronwall lemma we obtain

$$|v_t^{n, \varepsilon} - v_t^n|_E \leq c \int_0^t |F(v_s^n + z_s, u_s^\varepsilon) - F(v_s^n + z_s, u_s)|_E ds.$$

Letting  $n \rightarrow \infty$

$$|v_t^\varepsilon - v_t|_E \leq c \int_0^t |\delta^\varepsilon F_s|_E ds. \quad (32)$$

Now recall that  $\xi_t^\varepsilon := X_t^\varepsilon - X_t = v_t^\varepsilon - v_t$ , using the polynomial growth of  $F$  we end up with

$$|\xi_t^\varepsilon|_E \leq c \int_{E_\varepsilon} |\delta^\varepsilon F_s|_E ds \leq C_F(\omega)\varepsilon,$$

where  $C_F(\omega)$  has finite moments of any order. Hence  $\mathbb{E} \sup_t |\xi_t^\varepsilon|_E^{2k} \leq C_1 \varepsilon^{2k}$ ,  $k \geq 1$ .

(ii). Let  $Y^{n,\varepsilon}$  be the solution to the following equation

$$\frac{d}{dt} Y_t^{n,\varepsilon} = A_n Y_t^{n,\varepsilon} + D_x F(X_t, u_t) Y_t^{n,\varepsilon} + \delta^\varepsilon F_t. \quad (33)$$

Using the dissipativity assumptions we get

$$\begin{aligned} \frac{d^-}{dt} |Y_t^{n,\varepsilon}|_E &\leq \langle A_n Y_t^{n,\varepsilon} + D_x F(X_t, u_t) Y_t^{n,\varepsilon} + \delta^\varepsilon F_t, \delta_{Y_t^{n,\varepsilon}} \rangle_E \\ &\leq c |Y_t^{n,\varepsilon}|_E + |\delta^\varepsilon F_t|_E, \end{aligned}$$

then using the same strategy as before and passing to the limit with  $n \rightarrow \infty$  we get the required result.

(iii). Now let us define  $\eta_t^{n,\varepsilon} := v_t^{n,\varepsilon} - v_t^n - Y_t^{n,\varepsilon}$ . If  $\delta_t \in \partial |\eta_t^{n,\varepsilon}|$ , we have

$$\begin{aligned} \frac{d^-}{dt} |\eta_t^{n,\varepsilon}|_E &\leq \langle A_n \eta_t^{n,\varepsilon} + F(v_t^{n,\varepsilon} + z_t, u_t^\varepsilon) - F(v_t^n + z_t, u_t^\varepsilon) \\ &\quad - D_x F(v_t + z_t, u_t)(v_t^{n,\varepsilon} - v_t^n) + D_x F(v_t + z_t, u_t) \eta_t^{n,\varepsilon}, \delta_t \rangle_E \\ &\leq \langle \int_0^1 D_x F(v_t^n + z_t + \theta(v_t^{n,\varepsilon} - v_t^n), u_t^\varepsilon)(v_t^{n,\varepsilon} - v_t^n) d\theta \\ &\quad - D_x F(v_t + z_t, u_t)(v_t^{n,\varepsilon} - v_t^n), \delta_t \rangle_E + c |\eta_t^{n,\varepsilon}|_E \\ &= \langle \int_0^1 [D_x F(v_t^n + z_t + \theta(v_t^{n,\varepsilon} - v_t^n), u_t^\varepsilon) - D_x F(v_t^n + z_t, u_t^\varepsilon)](v_t^{n,\varepsilon} - v_t^n) d\theta, \delta_t \rangle_E \\ &\quad + \langle [D_x F(v_t^n + z_t, u_t^\varepsilon) - D_x F(v_t + z_t, u_t^\varepsilon)](v_t^{n,\varepsilon} - v_t^n), \delta_t \rangle_E \\ &\quad + \delta^\varepsilon D_x F_t \cdot (v_t^{n,\varepsilon} - v_t^n), \delta_t \rangle_E + c |\eta_t^{n,\varepsilon}|_E. \end{aligned}$$

Thanks to the Gronwall lemma we obtain

$$\begin{aligned} |\eta_t^{n,\varepsilon}|_E &\leq C \int_0^t \int_0^1 \left| D_x F(v_s^n + z_s + \theta(v_t^{n,\varepsilon} - v_t^n), u_s^\varepsilon) - D_x F(v_s^n + z_s, u_s^\varepsilon) \right|_{\mathcal{L}(E)} |v_s^{n,\varepsilon} - v_s^n|_E d\theta ds \\ &\quad + C \int_0^t \left| D_x F(v_s^n + z_s, u_s^\varepsilon) - D_x F(v_s + z_s, u_s^\varepsilon) \right|_{\mathcal{L}(E)} |v_s^{n,\varepsilon} - v_s^n|_E ds \\ &\quad + C \int_{E_\varepsilon} \left| \delta^\varepsilon D_x F_s \right|_{\mathcal{L}(E)} |v_s^{n,\varepsilon} - v_s^n|_E ds \end{aligned}$$

Letting  $n \rightarrow 0$

$$\begin{aligned} |\eta_t^\varepsilon|_E &\leq C \int_0^t \int_0^1 \left| D_x F(v_s + z_s + \theta \xi_s^\varepsilon, u_s^\varepsilon) - D_x F(v_s + z_s, u_s^\varepsilon) \right|_{\mathcal{L}(E)} |\xi_s^\varepsilon|_E d\theta ds \\ &\quad + C \int_{E_\varepsilon} \left| \delta^\varepsilon D_x F_s \right|_{\mathcal{L}(E)} |\xi_s^\varepsilon|_E ds, \end{aligned}$$

thanks to the continuity of the map  $DF(\cdot) : E \rightarrow \mathcal{L}(E)$ . Now according to the estimate obtained in (i) and to the Hölder inequality we obtain

$$\begin{aligned} \mathbb{E} \sup_{t \in [0, T]} |\eta_t^\varepsilon|_E^{2k} &\leq C \varepsilon^{2k} \cdot \left( \mathbb{E} \int_0^T \int_0^1 \left| D_x F(v_s + z_s + \theta \xi_s^\varepsilon, u_s^\varepsilon) - D_x F(v_s + z_s, u_s^\varepsilon) \right|_{\mathcal{L}(E)}^{4k} d\theta ds \right)^{1/2} \\ &\quad + C \varepsilon^{2k} \left( \mathbb{E} \int_{E_\varepsilon} \left| \delta^\varepsilon D_x F_s \right|_{\mathcal{L}(E)}^{4k} ds \right)^{1/2}. \end{aligned}$$

The continuity of the map  $DF(\cdot) : E \rightarrow \mathcal{L}(E)$  implies again that  $|D_x F(v_t + z_t + \theta \xi_t, u_t^\varepsilon) - D_x F(v_t + z_t, u_t^\varepsilon)|_{\mathcal{L}(E)}$  tends to zero if  $\varepsilon \rightarrow 0$ . Thanks to the polynomial growth of  $D_x F$  we get the result.  $\square$

Now we can deal with the expansion of the cost.

**Proposition 5.2.** *We have the following*

$$J(u^\varepsilon) - J(u) = \mathbb{E} \int_0^T [\delta^\varepsilon L_t + D_x L(t, X_t, u_t) Y_t^\varepsilon] dt + \mathbb{E} (D_x G(X_T) Y_T^\varepsilon) + o(\varepsilon),$$

where  $\delta^\varepsilon L_t = L(t, X_t, u_t^\varepsilon) - L(t, X_t, u_t)$ .

*Proof.* The difference between the two cost functionals reads as

$$J(u^\varepsilon) - J(u) = \mathbb{E} \int_0^T [L(t, X_t^\varepsilon, u_t^\varepsilon) - L(t, X_t, u_t)] dt + \mathbb{E} [G(X_T^\varepsilon) - G(X_T)]$$

Let us begin rewriting the running cost part

$$\begin{aligned} & \mathbb{E} \int_0^T [L(t, X_t^\varepsilon, u_t^\varepsilon) - L(t, X_t, u_t)] dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt \\ &= \mathbb{E} \int_0^T \int_0^1 D_x L(t, X_t + \theta \xi_t^\varepsilon, u_t^\varepsilon) \xi_t^\varepsilon d\theta dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt \\ &= \mathbb{E} \int_0^T \int_0^1 [D_x L(t, X_t + \theta \xi_t^\varepsilon, u_t^\varepsilon) - D_x L(t, X_t, u_t^\varepsilon)] d\theta \xi_t^\varepsilon dt \\ &+ \mathbb{E} \int_0^T D_x L(t, X_t, u_t^\varepsilon) \xi_t^\varepsilon dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt \\ &\leq C \left( \mathbb{E} \sup_{t \in [0, T]} |\xi_t^\varepsilon|_E^2 \right)^{1/2} \left( \mathbb{E} \left( \int_0^T \left| \int_0^1 [D_x L(t, X_t + \theta \xi_t^\varepsilon, u_t^\varepsilon) - D_x L(t, X_t, u_t^\varepsilon)] d\theta \right| dt \right)^2 \right)^{1/2} \\ &+ \mathbb{E} \int_0^T D_x L(t, X_t, u_t^\varepsilon) Y_t^\varepsilon dt + \mathbb{E} \int_0^T D_x L(t, X_t, u_t^\varepsilon) \eta_t^\varepsilon dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt \\ &\leq C \varepsilon \cdot \left( \mathbb{E} \left( \int_0^T \left| \int_0^1 [D_x L(t, X_t + \theta \xi_t^\varepsilon, u_t^\varepsilon) - D_x L(t, X_t, u_t^\varepsilon)] d\theta \right| dt \right)^2 \right)^{1/2} \\ &+ (\mathbb{E}(\sup_{t \in [0, T]} |\eta_t|_E^2))^{1/2} \left( \mathbb{E} \left( \int_0^T D_x L(t, X_t, u_t^\varepsilon) dt \right)^2 \right)^{1/2} \\ &+ \mathbb{E} \int_0^T D_x L(t, X_t, u_t^\varepsilon) Y_t^\varepsilon dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt \\ &= \mathbb{E} \int_0^T D_x L(t, X_t, u_t^\varepsilon) Y_t^\varepsilon dt + \mathbb{E} \int_0^T \delta^\varepsilon L_t dt + o(\varepsilon), \end{aligned}$$

where we used the estimates given in Lemma 5.1, the continuity of the map  $DL(\cdot) : E \rightarrow \mathcal{L}(E)$



and the polynomial growth. Regarding the second term we have

$$\begin{aligned}
\mathbb{E}[G(X_T^\varepsilon) - G(X_T)] &= \mathbb{E} \int_0^1 D_x G(X_T + \theta \xi_T^\varepsilon) d\theta \cdot \xi_T^\varepsilon \\
&= \mathbb{E} \int_0^1 [D_x G(X_T + \theta \xi_T^\varepsilon) - D_x G(X_T)] d\theta \cdot \xi_T^\varepsilon + \mathbb{E}(D_x G(X_T) \xi_T^\varepsilon) \\
&\leq C\varepsilon \cdot \left( \mathbb{E} \left( \int_0^1 [D_x G(X_T + \theta \xi_T^\varepsilon) - D_x G(X_T)] d\theta \right)^2 \right)^{1/2} \\
&\quad + \mathbb{E}(D_x G(X_T) Y_T^\varepsilon) + \mathbb{E} \sup_{t \in [0, T]} |\eta_t^\varepsilon|_E D_x G(X_t) \\
&= \mathbb{E}(D_x G(X_T) Y_T^\varepsilon) + o(\varepsilon),
\end{aligned}$$

where we used again the continuity of the map  $DG(\cdot) : E \rightarrow \mathcal{L}(E)$  along with Lemma 5.1. By adding the two terms we conclude the proof.  $\square$

**Remark 5.1.** In the particular case in which the cost is given by (26) the expansion of the cost reads as

$$\begin{aligned}
J(u^\varepsilon) - J(u) &= \mathbb{E} \int_0^T \int_{\mathcal{O}} [l(t, X_t(\xi), u_t^\varepsilon) - l(t, X_t(\xi), u_t) + D_x l(t, X_t(\xi), u_t) Y_t^\varepsilon(\xi)] \mu(d\xi) dt \\
&\quad + \mathbb{E} \int_{\mathcal{O}} Dg(X_T(\xi)) Y_T^\varepsilon(\xi) \mu(d\xi) + o(\varepsilon).
\end{aligned}$$

## 6 Proof of Theorem 3.1

The central idea in this proof is the construction of a pair of processes  $(p, q)$  by a duality argument. We introduce the following auxiliary equation:

$$\begin{cases} dy(t) = [Ay(t) + D_x F(X(t), u(t))y(t) + \gamma(t)]dt + \eta(t)dW(t) \\ y(0) = 0, \end{cases} \quad (34)$$

which is a generalized version of the first variation equation introduced before. We want to study the linear map

$$\tau : (\gamma(\cdot), \eta(\cdot)) \mapsto (y(\cdot), y(T)),$$

which assigns to the forcing terms  $\gamma(\cdot), \eta(\cdot)$  the solution  $y(\cdot)$  and its terminal value  $y(T)$ . For an appropriate choice of functional spaces,  $\tau$  turns out to be well defined and continuous, and so is its adjoint

$$\tau^* : (f(\cdot), \zeta) \mapsto (p(\cdot), q(\cdot)).$$

The pair we are looking for is defined as  $(p, q) = \tau^*(f, \zeta)$  when we choose  $\zeta = D_x H(X(T))^*$  and  $f(t) = D_x L(t, X(t), u(t))^*$ . The following Proposition will allow to identify the appropriate norms for this duality argument.

**Proposition 6.1.** *Let Hypotheses 2.1, 2.2 and 2.5 be in force. Suppose we are given  $r > 2$  and two processes  $\gamma, \eta$  such that  $\gamma \in L^r_{\mathcal{F}}(\Omega; L^2([0, T], -\lambda; H))$  and  $\eta \in L^r_{\mathcal{F}}(\Omega; L^2([0, T]; \mathcal{L}_2(K, V)))$ .*

1. *The equation (34) has a unique mild solution, i.e. a progressive  $E$ -valued process  $y$  such that,  $\mathbb{P}$ -a.s.,  $t \mapsto y(t)$  belongs to  $L^2([0, T]; E)$  and*

$$y(t) = \int_0^t S(t-s)(D_x F(X(s), u(s))y(s) + \gamma(s))ds + \int_0^t S(t-s)\eta(s)dW(s), \quad (35)$$

for almost every  $t \in [0, T]$ .

2. The following estimate hold:

$$\mathbb{E} \int_0^T |y(t)|_E^2 dt \leq C \left[ \mathbb{E} \left( \int_0^T |\gamma(s)|_H^2 (T-s)^{-\lambda} ds \right)^{r/2} + \mathbb{E} \left( \int_0^T \|\eta(s)\|_{L_2(K,V)}^2 ds \right)^{r/2} \right]^{\frac{2}{r}} \quad (36)$$

If the forcing term  $\gamma(\cdot)$  belongs to  $L^r(\Omega; C([0, T]; H))$  then the solution is bounded in  $L^2(\Omega; E)$ , more precisely the following estimate holds for every  $t \in [0, T]$ :

$$\mathbb{E} |y(t)|_E^2 \leq C \left[ \mathbb{E} \left( \int_0^t |\gamma(s)|_H^2 (t-s)^{-\lambda} ds \right)^{r/2} + \mathbb{E} \left( \int_0^t \|\eta(s)\|_{L_2(K,V)}^2 ds \right)^{r/2} \right]^{\frac{2}{r}} \quad (37)$$

*Proof. Existence and uniqueness.* Let us denote

$$T(t) := D_x F(X(t), u(t)), \quad \Gamma(t) := \int_0^t S(t-s) \gamma(s) ds, \quad W_\eta(t) := \int_0^t S(t-s) \eta(s) dW(s).$$

We note for further use that, for  $\mathbb{P}$ -almost every  $\omega \in \Omega$ ,  $\sup_{t \in [0, T]} \|T(t)\|_{\mathcal{L}(E)}$  is bounded by a constant  $C_F(\omega)$ . Then equation (35) can be written:  $\mathbb{P}$ -a.s.,

$$y(t) = \int_0^t S(t-s) T(s) y(s) ds + \Gamma(t) + W_\eta(t) \quad \text{for almost every } t \in [0, T]. \quad (38)$$

Now we want to prove that  $\Gamma$  and  $W_\eta$  belong to  $L^2_{\mathcal{F}}([0, T]; E)$ .

$$\begin{aligned} \mathbb{E} \int_0^T \left| \int_0^t S(t-s) \gamma(s) ds \right|_E^2 dt &\leq \mathbb{E} \int_0^T \left( \int_0^t |S(t-s) \gamma(s)|_E ds \right)^2 dt \\ &\leq C \mathbb{E} \int_0^T \left( \int_0^t |\gamma(s)|_H (t-s)^{-\lambda} ds \right)^2 dt \\ &\leq C \mathbb{E} \int_0^T \left( \int_0^t |\gamma(s)|_H^2 (t-s)^{-\lambda} ds \cdot \int_0^t (t-s)^{-\lambda} ds \right) dt \\ &\leq C \mathbb{E} \int_0^T \left( \int_0^t |\gamma(s)|_H^2 (t-s)^{-\lambda} ds \right) dt \\ &= C \mathbb{E} \int_0^T |\gamma(s)|_H^2 \left( \int_s^T (t-s)^{-\lambda} dt \right) ds \\ &= C \mathbb{E} \int_0^T |\gamma(s)|_H^2 (T-s)^{1-\lambda} ds < \infty \end{aligned} \quad (39)$$

thanks to the fact that  $\lambda < 1$  and  $\gamma \in L^r_{\mathcal{F}}(\Omega; L^2([0, T], -\lambda; H)) \subset L^2_{\mathcal{F}}([0, T]; H)$ . Regarding the stochastic convolution we have

$$\begin{aligned} \mathbb{E} \int_0^T \left| \int_0^t S(t-s) \eta(s) dW_s \right|_E^2 dt &\leq C \mathbb{E} \int_0^T \left| \int_0^t S(t-s) \eta(s) dW_s \right|_V^2 dt \\ &\leq C \mathbb{E} \int_0^T \|\eta(s)\|_{L_2(K,V)}^2 ds < \infty \end{aligned} \quad (40)$$

where we have used the Ito isometry for the stochastic integral in the Hilbert space  $V$  with respect to cylindrical noise, as well as Hypothesis 2.2. If we define

$$v(t) = y(t) - \Gamma(t) - W_\eta(t),$$

then equation (35) is equivalent to the following:  $\mathbb{P}$ -a.s.,

$$v(t) = \int_0^t S(t-s) T(s) v(s) ds + \int_0^t S(t-s) T(s) (\Gamma(s) + W_\eta(s)) ds \quad (41)$$

for almost every  $t \in [0, T]$ . Recalling that  $\sup_{t \in [0, T]} \|T(t)\|_{\mathcal{L}(E)} \leq C_F(\omega)$  we see that  $\mathbb{P}$ -a.s.  $s \mapsto T(s)(\Gamma(s) + W_\eta(s))$  belongs to  $L^2([0, T]; E)$  and then it can be proved by an easy contraction argument that there exists a unique pathwise solution  $v(\omega, \cdot) \in L^2([0, T]; E)$  (in fact,  $v(\omega, \cdot) \in C([0, T]; E)$ ) and that,  $\mathbb{P}$ -a.s., equality (41) holds for every  $t \in [0, T]$ .

By the same arguments, uniqueness of the pathwise solution  $v$  implies uniqueness for the original equation (35).

**Estimates.** Let us now define the Yosida approximations  $A_n = nA(n - A)^{-1}$  of  $A$  and consider the approximating equations

$$v'_n(t) = A_n v_n(t) + T(t)v_n(t) + T(t)(\Gamma(t) + W_\eta(t)), \quad v(0) = 0. \quad (42)$$

Then, for any  $t \in [0, T]$ , we have  $\mathbb{P}$ -a.s.

$$\frac{d^-}{dt} |v_n(t)|_E \leq \langle A_n v_n(t), \delta_n(t) \rangle_E + \langle T(t)(v_n(t) + \Gamma(t) + W_\eta(t)), \delta_n(t) \rangle_E \quad (43)$$

where  $\delta_n(t) \in \partial|v_n(t)|_E$ . Using the dissipativity properties proved in Lemma 2.7 we have  $\langle T(t)v_n(t), \delta_n(t) \rangle_E \leq cv_n(t)_E$  and from the contraction property of the semigroup it follows that  $\langle A_n v_n(t), \delta_n(t) \rangle_E \leq 0$ . So we obtain

$$\frac{d^-}{dt} |v_n(t)|_E \leq cv_n(t)_E + C_F(\omega)|(\Gamma(t) + W_\eta(t))|_E, \quad \mathbb{P}\text{-a.s.} \quad (44)$$

where the constant  $C_F(\omega)$  is finite for a.e.  $\omega \in \Omega$ . Using Gronwall's lemma

$$|v_n(t)|_E \leq C(\omega) \int_0^t |(\Gamma(s) + W_\eta(s))|_E ds, \quad \mathbb{P}\text{-a.s.} \quad (45)$$

Now, it is easy to prove that  $v_n(t) \rightarrow v(t)$  in  $E$  so we can take the limit as  $n$  goes to infinity and we obtain

$$|v(t)|_E \leq C(\omega) \int_0^t |(\Gamma(s) + W_\eta(s))|_E ds. \quad (46)$$

Recalling the definition of  $v(t)$  we get

$$\begin{aligned} |y(t)|_E &\leq C(\omega) \int_0^t |(\Gamma(s) + W_\eta(s))|_E ds + |\Gamma(t)|_E + |W_\eta(t)|_E \\ &\leq C(\omega) (|\Gamma(t)|_E + |W_\eta(t)|_E), \quad \mathbb{P}\text{-a.s.} \end{aligned} \quad (47)$$

To prove estimate (36) we have to consider

$$\begin{aligned} \mathbb{E} \int_0^T |y(t)|_E^2 dt &\leq \mathbb{E} \left[ C(\omega)^2 \int_0^T (|\Gamma(t)|_E + |W_\eta(t)|_E)^2 dt \right] \\ &\leq \left[ \mathbb{E}(C(\omega)^{2r'}) \right]^{1/r'} \left[ \mathbb{E} \left( \int_0^T (|\Gamma(t)|_E^2 + |W_\eta(t)|_E^2) dt \right)^r \right]^{1/r} \\ &\leq K \left[ \mathbb{E} \left( \int_0^T |\Gamma(t)|_E^2 + |W_\eta(t)|_E^2 dt \right)^r \right]^{1/r} \end{aligned}$$

where we chose  $r \geq 2$  and we used the finiteness of the moments of  $C(\omega)$ . Using the same computations as in (39) and (40) we end up with

$$\left( \mathbb{E} \int_0^T |y(t)|_E^2 dt \right)^{\frac{1}{2}} \leq C \left[ \mathbb{E} \left( \int_0^T |\gamma(s)|_H^2 (T-s)^{-\lambda} ds \right)^{r/2} + \mathbb{E} \left( \int_0^T \|\eta(s)\|_{L_2(K, V)}^2 ds \right)^{r/2} \right]^{\frac{1}{r}}$$

which is exactly what we looked for.

Assuming now that  $\gamma \in L^r(\Omega; C([0, T]; H))$  then we have

$$\begin{aligned} |\Gamma(t)|_E^2 &= \left| \int_0^t S(t-s)\gamma(s)ds \right|_E^2 \leq \left( \int_0^t |S(t-s)\gamma(s)|_E ds \right)^2 \\ &\leq C \left( \int_0^t |\gamma(s)|_H (t-s)^{-\lambda} ds \right)^2 \\ &\leq C \int_0^t |\gamma(s)|_H^2 (t-s)^{-\lambda} ds \cdot \int_0^t (t-s)^{-\lambda} ds \\ &\leq C \int_0^t |\gamma(s)|_H^2 (t-s)^{-\lambda} ds \end{aligned}$$

thanks to the fact that  $\lambda < 1$ . Regarding the stochastic convolution we have

$$\begin{aligned} \mathbb{E}|W_\eta(t)|_E^2 &= \mathbb{E} \left| \int_0^t S(t-s)\eta(s)dW(s) \right|_E^2 \leq C \cdot \mathbb{E} \left| \int_0^t S(t-s)\eta(s)dW(s) \right|_V^2 \\ &\leq C \cdot \mathbb{E} \int_0^t \|\eta(s)\|_{L_2(K,V)}^2 ds. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbb{E}|y(t)|_E^2 &\leq \mathbb{E} \left( C(\omega)^2 \left( |\Gamma(t)|_E^2 + |\eta(t)|_E^2 \right) \right) \\ &\leq \left[ \mathbb{E}(C(\omega)^{2r'}) \right]^{1/r'} \left[ \mathbb{E} \left( |\Gamma(t)|_E^2 + |W_\eta(t)|_E^2 \right)^r \right]^{1/r} \\ &\leq K \left[ \mathbb{E} \left( |\Gamma(t)|_E^2 + |W_\eta(t)|_E^2 \right)^r \right]^{1/r}. \end{aligned}$$

The inequality (37) now follows immediately.  $\square$

By Proposition 6.1,  $\tau$  is a bounded linear operator from the space  $L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda; H)) \times L_{\mathcal{F}}^r(\Omega; L^2([0, T]; L_2(K, V)))$  to  $L_{\mathcal{F}}^2(\Omega \times [0, T]; E) \times L_{\mathcal{F}_T}^2(\Omega; E)$ . Its dual operator  $\tau^*$  is also a bounded linear map

$$\tau^* : L_{\mathcal{F}}^2(\Omega \times [0, T]; E)' \times L_{\mathcal{F}_T}^2(\Omega; E)' \longrightarrow L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], -\lambda; H')) \times L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T]; L_2(K, V)))$$

where  $1/r + 1/r' = 1$ . We have used the fact that  $[L_{\mathcal{F}}^2(\Omega \times [0, T]; \lambda; H)]' = L_{\mathcal{F}}^2(\Omega \times [0, T], -\lambda; H')$  and that the dual of  $L_2(K, V)$  can be identified with  $L_2(K, V')$  (see e.g. [3], page 291).

**Remark 6.1.** If  $B$  is a separable and reflexive Banach space, then the dual of  $L_{\mathcal{F}}^2(\Omega \times [0, T]; B)$  is  $L_{\mathcal{F}}^2(\Omega \times [0, T]; B')$  (cfr. e.g. [11]). In our case,  $E$  is the space of real continuous functions, so this result no longer holds. However it is still true that  $L_{\mathcal{F}}^2(\Omega \times [0, T]; B') \subset L_{\mathcal{F}}^2(\Omega \times [0, T]; B)'$ . For our purposes, this is sufficient because we need only to evaluate  $\tau^*$  at  $f$  and  $\zeta$  which are much more regular.

**Remark 6.2.** A priori it could be possible to choose  $\gamma$  with values in  $E'$  instead of  $H$ . However this would introduce  $[L_{\mathcal{F}}^r(\Omega; L^2([0, T]; E))]'$ , which is difficult to treat.

We note that by definition the following equality holds

$$\mathbb{E} \int_0^T {}_{H'} \langle p(t), \gamma(t) \rangle_H dt + \mathbb{E} \int_0^T \langle q(t), \eta(t) \rangle_{L_2(K,V)} dt = \mathbb{E} \int_0^T \langle f(t), y(t) \rangle_E dt + \mathbb{E} \langle \zeta, y(T) \rangle_E, \quad (48)$$

where  $(p, q) = \tau^*(f, \zeta)$ , and  $(y(\cdot), y(T))$  are the solution process and its terminal value of equation (35) corresponding to forcing terms  $(\gamma, \eta)$ .

*Proof of Theorem 3.1.* Since  $u$  is an optimal control we have  $J(u^\varepsilon) - J(u) \geq 0$ . Thanks to the estimates given in Lemma 5.1, we already obtained in Proposition 5.2 that

$$o(\varepsilon) \leq \mathbb{E} \int_0^T [\delta^\varepsilon L(t) + D_x L(t, X(t), u(t)) Y_t^\varepsilon] dt + \mathbb{E} (D_x G(X_T) Y_T^\varepsilon(x)). \quad (49)$$

Now we use duality (more precisely we use equation (48) with  $\eta = 0$ ) for the first variation equation (28), which reads as

$$\mathbb{E} \langle D_x G(X(T)), Y^\varepsilon(T) \rangle_E = \mathbb{E} \int_0^T [{}_{H'} \langle p(t), \delta^\varepsilon F(t) \rangle_H - \langle D_x L(t, X(t), u(t)), Y^\varepsilon(t) \rangle_E] dt. \quad (50)$$

Here we have to check that  $\delta^\varepsilon F \in L^r_{\mathcal{F}}(\Omega; L^2([0, T], -\lambda; H))$  but this is true thanks to Hypothesis 2.5. Substituting in (49) we get

$$o(1) \leq \frac{1}{\varepsilon} \mathbb{E} \int_{t_0}^{t_0+\varepsilon} [\mathcal{H}(t, X(t), u^\varepsilon(t), p(t)) - \mathcal{H}(t, X(t), u(t), p(t))] dt. \quad (51)$$

Now the proof can be concluded by usual arguments, see for instance [26]. □

## 7 The adjoint equation and the proof of Theorem 3.2

In this section we study a backward stochastic partial differential equation (BSPDE) and we characterize the adjoint processes  $(p, q)$  as its unique solution. This leads immediately to the proof of Theorem 3.2. While the process  $p$  takes values in  $E'$ , it is not easy to formulate the BSPDE as an equation in this space, since the usual tools of stochastic calculus are not available there, in particular there is no version of the martingale representation theorem in  $E'$ . We will therefore embed  $E'$  in a bigger Hilbert space in which we can use many standard techniques and we can uniquely solve the BSPDE in a mild formulation. Then we will prove that the solution is indeed more regular and in particular that  $p$  takes values in  $E'$  (and even in  $H'$ ) as desired.

Recall that if we identify  $H \simeq H'$ , we can embed the Gelfand triple  $E \subset H' \subset E'$  in a Hilbertian triple  $V \subset H' \subset V'$  and we obtain the following dense continuous inclusions

$$V \subset E \subset H \simeq H' \subset E' \subset V',$$

where now we suppose that Hypothesis 2.3 holds true. We first write the dual BSPDE in a formal way as follows

$$\begin{cases} -dp(t) = [A'p(t) + D_x F(X(t), u(t))' p(t) + f(t)] dt - q(t) dW(t) \\ p(T) = \zeta, \end{cases} \quad (52)$$

where  $f(t) := D_x L(t, X(t), u(t))^*$  and  $\zeta := D_x G(X(T))^*$  take values in  $E'$ . This equation is given the following precise meaning in a mild formulation: for all  $t \in [0, T]$ ,  $\mathbb{P}$ -a.s.

$$\begin{aligned} p(t) = & S(T-t)' \zeta + \int_t^T S(s-t)' [D_x F(X(s), u(s))^* p(s) + f(s)] ds \\ & - \int_t^T S(s-t)^\sim q(s) dW(s), \end{aligned} \quad (53)$$

where  $S(t)'\colon E' \rightarrow H'$  is the adjoint of  $S(t)\colon H \rightarrow E$  (compare Hypothesis 2.1-1) and  $S(t)^\sim\colon V' \rightarrow V'$  is the adjoint of  $S(t)$  viewed as an operator from  $V$  to  $V$ , which is possible thanks to Hypothesis 2.2. On the other hand,  $T^*\colon E' \rightarrow E'$  denotes the dual of any bounded linear operator  $T\colon E \rightarrow E$ . Note that equality (53) has a meaning in the space  $V'$ , in which the stochastic convolution takes values.

**Remark 7.1.** Notice that, with the previous notation, the adjoint of  $S(t) : H \rightarrow H$  (i.e.,  $S(t)$  viewed as an operator on  $H$  rather than from  $H$  to  $E$ ) coincides with  $S(t)'|_{H'} : H' \rightarrow H'$ , the restriction of  $S(t)' : E' \rightarrow H'$  to  $H' \subset E'$ .

**Remark 7.2.** Notice that the BSPDE is linear in  $p$ , but the map  $(t, \omega) \mapsto D_x F(X(t), u(t))^* : E' \rightarrow E'$  is not bounded. Using the Nemytskii characterization it is easy to see that  $D_x F(x, u)'$  acts on measures as multiplication by a density. Indeed, for any  $v \in E$

$$\begin{aligned} \langle D_x F(x, u)^* p, v \rangle_E &= \langle p, D_x F(x, u) v \rangle_E \\ &= \int_{\mathcal{O}} f'(x(\xi), u) v(\xi) dp(\xi) \\ &= \int_{\mathcal{O}} v(\xi) \cdot f'(x(\xi), u) dp(\xi) \end{aligned} \quad (54)$$

which means that it maps  $p$  into  $f'(x, u)p$ . It is worth noting that also the terminal condition lives in  $E'$ , as well as the forcing term  $D_x L(t, X(t), u(t))^*$ .

Now we can state the following

**Theorem 7.1.** *Let Hypotheses 2.1, 2.3, 2.4, 2.5 hold and  $r' \in (1, 2)$ . Then there exists a unique mild solution  $(p, q) \in L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \lambda; H')) \times L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T]; \mathcal{L}_2(K, V)))$  to the BSPDE (53).*

*Proof. Existence.*

STEP 1: (Regularization) We shall construct a solution to equation (53) by means of an approximation technique. The main tool we use is the smoothing effect of the semigroup  $S(t)' : E' \rightarrow H'$ . If we define

$$\zeta^\varepsilon := S(\varepsilon)' \zeta, \quad f^\varepsilon(t) := S(\varepsilon)' f(t),$$

then both  $\zeta^\varepsilon, f^\varepsilon(t)$  take values in  $H'$ , for all  $\varepsilon > 0$ . We also introduce the Yosida approximations  $f_\alpha$  of  $f$ . Thanks to Lemma 2.8 we know that the associated Nemytskii operator  $F_\alpha(x, u)$  is Lipschitz in  $x$  with respect to the norm of  $H$ , hence  $\|D_x F_\alpha\|_{\mathcal{L}(H)}$  is uniformly bounded by some constant  $k_\alpha$  which depends only on  $\alpha$ . If we set  $T_\alpha(t) := D_x F_\alpha(X(t), u(t))$  then its adjoint  $T_\alpha(t)^*$  is a bounded process with values in  $\mathcal{L}(H')$ . Next we introduce an approximate BSPDE that we first write in a formal way as

$$\begin{cases} -dp^{\varepsilon, \alpha}(t) = [A' p^{\varepsilon, \alpha}(t) + T_\alpha(t)^* p^{\varepsilon, \alpha}(t) + f^\varepsilon(t)] dt - q^{\varepsilon, \alpha}(t) dW(t), \\ p^{\varepsilon, \alpha}(T) = \zeta^\varepsilon. \end{cases} \quad (55)$$

By the result in [18] there exists a unique mild solution to this equation, i.e. a process

$$(p^{\varepsilon, \alpha}, q^{\varepsilon, \alpha}) \in L_{\mathcal{F}}^2(\Omega \times [0, T], H') \times L_{\mathcal{F}}^2([0, T], \mathcal{L}_2(K, H')),$$

such that for all  $t > 0$ ,  $\mathbb{P}$ -a.s

$$p^{\varepsilon, \alpha}(t) = S(T-t)' \zeta^\varepsilon + \int_t^T S(s-t)' [T_\alpha(s)^* p^{\varepsilon, \alpha}(s) + f^\varepsilon(s)] ds + \int_t^T S(s-t)' q^{\varepsilon, \alpha}(s) dW(s). \quad (56)$$

Our aim now is to prove a uniform estimate for the approximate solution via a duality argument.

STEP 2: (Duality) For all  $\gamma \in L_{\mathcal{F}}^r(\Omega; C([0, T], H))$  and  $\eta \in L_{\mathcal{F}}^r(\Omega; L^2([0, T], \mathcal{L}_2(K, V)))$  let us consider the equation

$$y^\alpha(t) = \int_0^t S(t-s) [T_\alpha(s) y^\alpha(s) + \gamma(s)] ds + \int_0^t S(t-s) \eta(s) dW(s), \quad t \in [0, T],$$

which has a unique mild solution  $y^\alpha(t)$  in the sense of Proposition 6.1. For all  $n \in \mathbb{N}$ , let  $A_n = nA(n - A)^{-1}$  denote the Yosida approximations of the operator  $A$  and  $S_n(t) := e^{tA_n}$ . Then, in a similar way, we can find a solution to

$$y_n^\alpha(t) = \int_0^t S_n(t-s)[T_\alpha(s)y_n^\alpha(s) + \gamma(s)]ds + \int_0^t S_n(t-s)\eta(s)dW(s).$$

Again by [18] there exists a unique solution to the equation

$$p_n^{\varepsilon,\alpha}(t) = S_n(T-t)\zeta^\varepsilon + \int_t^T S_n(T-t)'[T_\alpha(s)^*p_n^{\varepsilon,\alpha}(s) + f^\varepsilon(s)]ds + \int_t^T S_n(T-t)'q^{\varepsilon,\alpha}(s)dW(s),$$

in the space  $L_{\mathcal{F}}^2(\Omega \times [0, T], H') \times L_{\mathcal{F}}^2([0, T], \mathcal{L}_2(K, H'))$ . Notice that, since  $A_n$  are bounded operators, the equations can now be written in the stronger form

$$\begin{cases} dy_n^\alpha(t) = [A_n y_n^\alpha(t) + T_\alpha(t)y_n^\alpha(t) + \gamma(t)]dt + \eta(t)dW(t), \\ y_n^\alpha(0) = 0; \\ -dp_n^{\varepsilon,\alpha}(t) = [A_n^* p_n^{\varepsilon,\alpha}(t) + T_\alpha(t)^* p_n^{\varepsilon,\alpha}(t) + f^\varepsilon(t)]dt - q^{\varepsilon,\alpha}(t)dW(t), \\ p_n^{\varepsilon,\alpha}(T) = \zeta^\varepsilon, \end{cases}$$

where  $A_n^* : H' \rightarrow H'$  is understood as the adjoint of  $A_n : H \rightarrow H$ . Computing the Ito differential  $d_H \langle y_n^\alpha(t), p_n^{\varepsilon,\alpha}(t) \rangle_{H'}$  and letting  $n \rightarrow \infty$  we get by standard arguments (see e.g. [25]):

$$\begin{aligned} & \mathbb{E} \int_0^T \langle p^{\varepsilon,\alpha}(t), \gamma(t) \rangle_H dt + \mathbb{E} \int_0^T \langle q^{\varepsilon,\alpha}(t), \eta(t) \rangle_{\mathcal{L}_2(K, H)} dt \\ &= \mathbb{E} \int_0^T \langle f^\varepsilon(t), y^\alpha(t) \rangle_H dt + \mathbb{E}_{H'} \langle \zeta^\varepsilon, y^\alpha(T) \rangle_H. \end{aligned} \quad (57)$$

Since, by Proposition 6.1,  $y^\alpha(t)$  takes values in  $E$  we also have

$$\begin{aligned} & \mathbb{E} \int_0^T \langle p^{\varepsilon,\alpha}(t), \gamma(t) \rangle_H dt + \mathbb{E} \int_0^T \langle q^{\varepsilon,\alpha}(t), \eta(t) \rangle_{\mathcal{L}_2(K, H)} dt \\ &= \mathbb{E} \int_0^T \langle f^\varepsilon(t), y^\alpha(t) \rangle_E dt + \mathbb{E}_{E'} \langle \zeta^\varepsilon, y^\alpha(T) \rangle_E. \end{aligned} \quad (58)$$

Let us now define the set  $\mathcal{A} := \{\gamma \in L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda, H)) : \|\gamma\|_{L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda, H))} \leq 1\}$  and  $\tilde{\mathcal{A}} := \{\gamma \in L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda, H)) \cap L_{\mathcal{F}}^r(\Omega; C([0, T], H)) : \|\gamma\|_{L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda, H))} \leq 1, \}$ .

It easy to see that  $\tilde{\mathcal{A}}$  is densely embedded in  $\mathcal{A}$ . If we take  $\eta = 0$  in (58) we get

$$\begin{aligned} & \left( \mathbb{E} \left( \int_0^T |p^{\varepsilon,\alpha}(t)|_{H'}^2 (T-t)^\lambda dt \right)^{r'/2} \right)^{1/r'} \\ & \leq \sup_{\gamma \in \mathcal{A}} \left[ \mathbb{E} \int_0^T \langle f^\varepsilon(t), y^\alpha(t) \rangle_E dt + \mathbb{E} \langle \zeta^\varepsilon, y^\alpha(T) \rangle_E \right] \\ & = \sup_{\gamma \in \tilde{\mathcal{A}}} \left[ \mathbb{E} \int_0^T \langle f^\varepsilon(t), y^\alpha(t) \rangle_E dt + \mathbb{E} \langle \zeta^\varepsilon, y^\alpha(T) \rangle_E \right] \\ & \leq \sup_{\gamma \in \tilde{\mathcal{A}}} \left[ \left( \mathbb{E} \int_0^T |f^\varepsilon(t)|_{E'}^2 dt \right)^{1/2} \left( \mathbb{E} \int_0^T |y^\alpha(t)|_E^2 dt \right)^{1/2} \right] \\ & \quad + \sup_{\gamma \in \tilde{\mathcal{A}}} \left[ \left( \mathbb{E} |\zeta^\varepsilon|_{E'}^2 \right)^{1/2} \left( \mathbb{E} |y^\alpha(T)|_E^2 \right)^{1/2} \right] \\ & \leq C \left[ \left( \mathbb{E} \int_0^T |f^\varepsilon(t)|_{E'}^2 dt \right)^{1/2} + \left( \mathbb{E} |\zeta^\varepsilon|_{E'}^2 \right)^{1/2} \right] \end{aligned}$$

where we used Hölder's inequality and the estimates (36) and (37) with  $t = T$  of Proposition 6.1 for the process  $y^\alpha(t)$ . Then, noting that  $|f^\varepsilon|_{E'} \leq |f|_{E'}$  and  $|\zeta^\varepsilon|_{E'} \leq |\zeta|_{E'}$ , we obtain

$$\left( \mathbb{E} \left( \int_0^T |p^{\varepsilon, \alpha}(t)|_{H'}^2 (T-t)^\lambda dt \right)^{r'/2} \right)^{1/r'} \leq C \left[ \mathbb{E} \int_0^T |f(t)|_{E'}^2 dt + \mathbb{E} |\zeta|_{E'}^2 \right]^{1/2}$$

where  $C$  does not depend on  $\varepsilon$  and  $\alpha$ . Let us now look for a similar estimate on  $q^{\varepsilon, \alpha}(t)$ . We first recall that if  $(e_k)_{k \in \mathbb{N}}$  is a complete orthonormal system of  $H$  such that  $e_k \in V$  for all  $k \in \mathbb{N}$ , we have

$$\begin{aligned} \langle q^{\varepsilon, \alpha}(t), \eta(t) \rangle_{\mathcal{L}_2(K, H)} &= \sum_k \langle q^{\varepsilon, \alpha}(t) e_k, \eta(t) e_k \rangle_H \\ &= \sum_k \langle q^{\varepsilon, \alpha}(t) e_k, \eta(t) e_k \rangle_V \\ &= \mathcal{L}_2(K, V') \langle q^{\varepsilon, \alpha}(t), \eta(t) \rangle_{\mathcal{L}_2(K, V)}. \end{aligned} \quad (59)$$

Hence we can identify  $\mathcal{L}_2(K, V)'$  with the space  $\mathcal{L}_2(K, V')$ : for a detailed proof see [3], page 291. Writing now (58) with  $\gamma = 0$ , letting  $\mathcal{B} := \{\eta \in L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \mathcal{L}_2(K, V))) : \|\eta\|_{L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \mathcal{L}_2(K, V)))} \leq 1\}$  and recalling the estimate (36) and (37) with  $t = T$  we get

$$\begin{aligned} &\left( \mathbb{E} \left( \int_0^T |q^{\varepsilon, \alpha}(t)|_{\mathcal{L}_2(K, V')}^2 dt \right)^{r'/2} \right)^{1/r'} \\ &\leq \sup_{\eta \in \mathcal{B}} \left[ \mathbb{E} \int_0^T \langle f^\varepsilon(t), y^\alpha(t) \rangle_E dt + \mathbb{E}_{E'} \langle \zeta^\varepsilon, y^\alpha(T) \rangle_E \right] \\ &\leq \sup_{\eta \in \mathcal{B}} \left[ \left( \mathbb{E} \int_0^T |f^\varepsilon(t)|_{E'}^2 dt \right)^{1/2} \left( \mathbb{E} \int_0^T |y^\alpha(t)|_E^2 dt \right)^{1/2} \right] \\ &\quad + \sup_{\eta \in \mathcal{B}} \left[ \left( \mathbb{E} |\zeta^\varepsilon|_{E'}^2 \right)^{1/2} \left( \mathbb{E} |y^\alpha(T)|_E^2 \right)^{1/2} \right] \\ &\leq C \left[ \left( \mathbb{E} \int_0^T |f^\varepsilon(t)|_{E'}^2 dt \right)^{1/2} + \left( \mathbb{E} |\zeta^\varepsilon|_{E'}^2 \right)^{1/2} \right] \end{aligned}$$

Summarizing, we have obtained the following estimate

$$\|p^{\varepsilon, \alpha}\|_{L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \lambda, H'))} + \|q^{\varepsilon, \alpha}\|_{L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \mathcal{L}_2(K, V')))} \leq C \left[ \mathbb{E} \int_0^T |f(t)|_{E'}^2 dt + \mathbb{E} |\zeta|_{E'}^2 \right]. \quad (60)$$

**STEP 3: (Convergence)** By (60) there exists a sequence  $(\varepsilon_k, \alpha_k) \rightarrow (0, 0)$ , as  $k \rightarrow \infty$ , such that  $(p^{\varepsilon_k, \alpha_k}, q^{\varepsilon_k, \alpha_k}) =: (p_k, q_k)$  converges to a limit  $(p, q)$  weakly in the product space. Setting for brevity  $T(t) = D_x F(X(t), u(t))$ , we will show that  $(p, q)$  is a mild solution to equation (53), which we now write in the form

$$p(t) = S(T-t)' \zeta + \int_t^T S(s-t)' [T(t)^* p(s) + f(s)] ds - \int_t^T S(s-t)' q(s) dW(s). \quad (61)$$

First we have  $S(T-t)' \zeta^{\varepsilon_k} \rightarrow S(T-t)' \zeta$  weakly in  $H'$  since, for any  $h \in H$ ,

$$_{H'} \langle S(T-t)' \zeta^{\varepsilon_k}, h \rangle_H = _{E'} \langle \zeta, S(\varepsilon_k) S(T-t) h \rangle_E \rightarrow _{E'} \langle \zeta, S(T-t) h \rangle_E = _{H'} \langle S(T-t)' \zeta, h \rangle_H.$$



Then we address the drift term. We take  $h \in H$  and we have

$$\begin{aligned} \mathbb{E} \int_t^T \int_{H'} \langle S(s-t)' T_{\alpha_k}(s)^* p_k(s), h \rangle_H ds &= \mathbb{E} \int_t^T \int_{H'} \langle T_{\alpha_k}(s)^* p_k(s), S(s-t)h \rangle_H ds \\ &= \mathbb{E} \int_t^T \int_{H'} \langle p_k(s), [T_{\alpha_k}(s) - T(s)] S(s-t)h \rangle_H ds \\ &\quad + \mathbb{E} \int_t^T \int_{H'} \langle p_k(s), T(s) S(s-t)h \rangle_H ds. \end{aligned}$$

Let us consider the first term on the right hand side:

$$\begin{aligned} &\mathbb{E} \int_t^T \int_{H'} \langle p_k(s), [T_{\alpha_k}(s) - T(s)] S(s-t)h \rangle_H ds \\ &= \mathbb{E} \int_t^T \langle (T-s)^{\lambda/2} p_k(s), (T-s)^{-\lambda/2} [T_{\alpha_k}(s) - T(s)] S(s-t)h \rangle_H ds \\ &\leq \left( \mathbb{E} \left( \int_t^T (T-s)^\lambda |p_k(s)|^2 ds \right)^{r'/2} \right)^{1/r'} \\ &\quad \cdot \left( \mathbb{E} \left( \int_t^T (T-s)^{-\lambda} |[T_{\alpha_k}(s) - T(s)] S(s-t)h|_H^2 ds \right)^{r/2} \right)^{1/r} \\ &\leq K \cdot \left( \mathbb{E} \left( \int_t^T (T-s)^{-\lambda} |[T_{\alpha_k}(s) - T(s)] S(s-t)h|_H^2 ds \right)^{r/2} \right)^{1/r}. \end{aligned} \tag{62}$$

Since  $S(s-t)h \in E$  (for  $s \neq t$ ) then  $|T_{\alpha_k}(s)S(s-t)h - T(s)S(s-t)h|_E \rightarrow 0$  for all  $\omega \in \Omega$ ,  $s \in [0, T]$ . Next we have

$$\begin{aligned} |T(s)S(s-t)h|_H^2 &= \int_{\mathcal{O}} |f'(X(s, \xi), u(s)) [S(s-t)h](\xi)|^2 m(dx) \\ &\leq \left( \sup_{\xi \in \mathcal{O}} |f'(X(s, \xi), u(s))| \right)^2 |S(s-t)h|_H^2 \\ &\leq C_F(\omega) \cdot |h|^2. \end{aligned} \tag{63}$$

where we have used the polynomial growth condition (6) on  $f'$  and  $C_F(\omega)$  was introduced in (30). Using (9) instead of (6) we prove by similar passages that  $|T_{\alpha_k}(s)S(s-t)h|_H^2 \leq C_F(\omega) \cdot |h|^2$ . Thanks to the assumption that  $\lambda < 1$  the right-hand side of (62) then tends to zero by dominated convergence.

To prove that  $\mathbb{E} \int_t^T \int_{H'} \langle p_k(s), T(s)S(s-t)h \rangle_H ds \rightarrow \mathbb{E} \int_t^T \int_{H'} \langle p(s), T(s)S(s-t)h \rangle_H ds$  it is enough to notice that  $s \rightarrow T(s)S(s-t)h 1_{s \geq t}$  lies in the space  $L_{\mathcal{F}}^r(\Omega; L^2([0, T], -\lambda, H))$  and to use the weak convergence property of  $p_k$ .

Finally, regarding the stochastic convolution it is enough to notice that the map

$$q \mapsto \int_t^T S(s-t) \sim q(s) dW(s)$$

is linear and continuous from  $L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T]; \mathcal{L}_2(K, V')))$  to  $L^{r'}(\Omega, V')$ , hence weakly continuous. Indeed, for fixed  $t$ , let us define:

$$I(\bar{q}) := \int_0^T (S(s-t) \sim 1_{\{s > t\}}) \bar{q}(s) dW(s)$$

Then, using Burkholder-Davis-Gundy inequality we get

$$\begin{aligned} \left( \mathbb{E} |I(\bar{q})|^{r'} \right)^{1/r'} &\leq K \left( \mathbb{E} \left( \int_0^T |S(s-t)' \bar{q}(s)|_{\mathcal{L}_2(K, V')}^2 ds \right)^{r'/2} \right)^{1/r'} \\ &\leq K \cdot \|\bar{q}\|_{L^{r'}(\Omega; L^2([0, T]; \mathcal{L}_2(K, V')))} < \infty, \end{aligned}$$

thanks to Hypothesis 2.3. Then the existence part follows.

### Uniqueness.

The BSPDE (53) is linear, then it is enough to prove that if the forcing term  $f(t)$  and the final condition  $\zeta(t)$  are zero then also the solution is identically zero. Hence we are dealing with the following

$$p(t) = \int_t^T S(s-t)' T(s)^* p(s) ds - \int_t^T S(s-t)' q(s) dW(s). \quad (64)$$

We apply the operator  $S(1/n)'$  to both sides and setting  $p^n := S(1/n)'p$ ,  $q^n := S(1/n)'q$  we obtain

$$p^n(t) = \int_t^T S(s-t)' S(1/n)' T(s)^* p(s) ds - \int_t^T S(s-t)' q^n(s) dW(s). \quad (65)$$

Notice that

$$\begin{aligned} |S(1/n)' T(s)^* p(s)|_{H'} &\leq |S(1/n)'|_{\mathcal{L}(E', H')} |T(s)^* p(s)|_{E'} \\ &\leq |S(1/n)'|_{\mathcal{L}(E', H')} |T(s)|_{\mathcal{L}(E)} |p(s)|_{E'} \\ &\leq C_n (1 + |X_s|_E^k) |p(s)|_{H'} \\ &\leq C_n(\omega) |p(s)|_{H'}, \quad \mathbb{P} - \text{a.s.} \end{aligned}$$

Moreover we have the following integrability condition: for  $\varepsilon > 9$ ,

$$\int_0^{T-\varepsilon} |p(s)|_{H'} dt < \infty, \quad \mathbb{P} - \text{a.s.} \quad (66)$$

thanks to the fact that  $p \in L_{\mathcal{F}}^{r'}(\Omega; L^2([0, T], \lambda; H'))$ . Notice that (66) fails in general for  $\varepsilon = 0$ . Next, by the semigroup property,  $S(1/n) = S(1/2n)S(1/2n)$ . By Hypothesis 2.1-1 we have  $S(1/2n)(H) \subset E$ . Now we exploit Hypothesis 2.3 to deduce that  $S(1/2n)(E) \subset V$  and we conclude that  $S(1/n)(H) \subset V$  and even that  $S(1/n) \in \mathcal{L}(H, V)$ , by the closed graph theorem. It follows that  $q^n(s) \in \mathcal{L}_2(K, H')$  and we have the estimate

$$|q^n(s)|_{\mathcal{L}_2(K, H')} \leq |S(1/n)'|_{\mathcal{L}(V', H')} |q(s)|_{\mathcal{L}_2(K, V')} = C_n |q(s)|_{\mathcal{L}_2(K, V')}, \quad (67)$$

which guarantees that the stochastic convolution in (65) is well defined. We write equation (65) on the time interval  $[0, T - \varepsilon]$  obtaining

$$p^n(t) = S(T - \varepsilon - t)' \zeta_\varepsilon^n + \int_t^{T-\varepsilon} S(s-t)' S(1/n)' T(s)^* p(s) ds - \int_t^{T-\varepsilon} S(s-t)' q^n(s) dW(s),$$

where  $\zeta_\varepsilon^n := p^n(T - \varepsilon)$  is the value of the solution at time  $T - \varepsilon$ . Next we fix an arbitrary  $\gamma$  in the space  $L_{\mathcal{F}}^2(\Omega \times [0, T], E)$  and define

$$y(t) = \int_0^t S(t-s) [T(s)y(s) + \gamma(s)] ds.$$

Note that by Proposition 6.1  $y$  is the solution to the equation (35) with  $\eta = 0$ .

Replacing the operator  $A$  with the Yosida approximations  $A_m$  we call  $p^{m,n}, y^m$  the solutions to the corresponding equations. We are in a position to apply the Itô formula to  $\langle p^{m,n}, y^m \rangle$ , thanks to the integrability conditions obtained in (66) and (67). Passing to the limit as  $m \rightarrow \infty$  and taking the expectation we obtain, see [25] for the details,

$$\begin{aligned} & \mathbb{E} \int_0^{T-\varepsilon} {}_{H'} \langle p^n(t), \gamma(t) \rangle_H dt + \mathbb{E} \int_0^{T-\varepsilon} [{}_{H'} \langle p^n, T(t)y(t) \rangle_H - {}_{H'} \langle S(1/n)'T(t)^*p(t), y(t) \rangle_H] dt \\ &= \mathbb{E}_{{}_{H'}} \langle y(T-\varepsilon), \zeta_\varepsilon^n \rangle_H. \end{aligned} \quad (68)$$

Now we pass to the limit letting  $n \rightarrow \infty$ . We have

$$\begin{aligned} & \mathbb{E} \int_0^{T-\varepsilon} {}_{H'} \langle p^n(t), \gamma(t) \rangle_H dt \rightarrow \mathbb{E} \int_0^{T-\varepsilon} {}_{H'} \langle p(t), \gamma(t) \rangle_H dt, \\ & \mathbb{E} \int_0^{T-\varepsilon} {}_{H'} \langle p^n(t), T(t)y(t) \rangle_H dt \rightarrow \mathbb{E} \int_0^{T-\varepsilon} {}_{H'} \langle p(t), T(t)y(t) \rangle_H dt, \end{aligned} \quad (69)$$

by dominated convergence, due to the fact that  $p(t)$  takes values in  $H'$ . We also have

$$\mathbb{E} \int_0^{T-\varepsilon} {}_{E'} \langle T(t)^*p(t), y(t) \rangle_E dt \rightarrow \mathbb{E} \int_0^{T-\varepsilon} {}_{E'} \langle T(t)^*p(t), S(1/n)y(t) \rangle_E dt$$

Indeed, from the analyticity of the semigroup in  $E$ , see [20], it follows that almost all paths of the process  $y$  take values in  $\overline{D(A)}$  (the closure of the domain of  $A$  in  $E$ ), so that  $S(1/n)y(t) \rightarrow y(t)$  in  $E$  and the desired conclusion follows from the dominated convergence theorem. So from (68), letting  $n \rightarrow \infty$ , it follows that

$$\mathbb{E} \int_0^{T-\varepsilon} \langle p(t), \gamma(t) \rangle dt = \mathbb{E} \langle y(T-\varepsilon), p(T-\varepsilon) \rangle. \quad (70)$$

Setting  $t = T - \varepsilon$  in (64) and taking the conditional expectation given  $\mathcal{F}_{T-\varepsilon}$  we obtain  $p(T-\varepsilon) = \mathbb{E} [\int_{T-\varepsilon}^T T(s)^*p(s)ds \mid \mathcal{F}_{T-\varepsilon}]$  so that

$$\begin{aligned} \mathbb{E} \int_0^{T-\varepsilon} \langle p(t), \gamma(t) \rangle dt &= \mathbb{E} \langle y(T-\varepsilon), \mathbb{E} [\int_{T-\varepsilon}^T T(s)^*p(s)ds \mid \mathcal{F}_{T-\varepsilon}] \rangle \\ &= \mathbb{E} \int_{T-\varepsilon}^T \langle y(T-\varepsilon), T(s)^*p(s) \rangle ds \end{aligned} \quad (71)$$

which converges to 0 if  $\varepsilon \rightarrow 0$ . We finally obtain  $\mathbb{E} \int_0^T \langle p(t), \gamma(t) \rangle dt = 0$  and we get the result by the arbitrariness of  $\gamma$ .  $\square$

We immediately arrive at the following conclusion.

*Proof of Theorem 3.2.* Thanks to Theorem 7.1, there exist a unique process  $p(\cdot)$  which is the first component of the solution to the BSPDE (53). The conclusion follows using Theorem 3.1.  $\square$

## References

- [1] Shmuel Agmon. On kernels, eigenvalues, and eigenfunctions of operators related to elliptic problems. *Comm. Pure Appl. Math.*, 18:627–663, 1965.
- [2] Antonio Ambrosetti and Giovanni Prodi. *A primer of nonlinear analysis*, volume 34 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1993.

- [3] Jean-Pierre Aubin. *Applied functional analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience, New York, second edition, 2000. With exercises by Bernard Cornet and Jean-Michel Lasry, Translated from the French by Carole Labrousse.
- [4] A. Bensoussan. Stochastic maximum principle for distributed parameter systems. *J. Franklin Inst.*, 315(5-6):387–406, 1983.
- [5] Stefano Bonaccorsi and Marco Fuhrman. Integration by parts and smoothness of the law for a class of stochastic evolution equations. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 7(1):89–129, 2004.
- [6] Zdzisław Brzeźniak and Rafael Serrano. Optimal relaxed control of dissipative stochastic partial differential equations in Banach spaces. *SIAM J. Control Optim.*, 51(3):2664–2703, 2013.
- [7] Sandra Cerrai. *Second order PDE’s in finite and infinite dimension*, volume 1762 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001. A probabilistic approach.
- [8] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.
- [9] Giuseppe Da Prato. *Kolmogorov equations for stochastic PDEs*. Springer, 2004.
- [10] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.
- [11] J. Diestel and J. J. Uhl, Jr. *Vector measures*. American Mathematical Society, Providence, R.I., 1977. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15.
- [12] Kai Du and Qingxin Meng. Stochastic maximum principle for infinite dimensional control systems. *preprint arXiv:1208.0529*, 2012.
- [13] Kai Du and Qingxin Meng. A maximum principle for optimal control of stochastic evolution equations. *SIAM J. Control Optim.*, 51(6):4343–4362, 2013.
- [14] Marco Fuhrman, Ying Hu, and Gianmario Tessitore. Stochastic maximum principle for optimal control of SPDEs. *C. R. Math. Acad. Sci. Paris*, 350(13-14):683–688, 2012.
- [15] Marco Fuhrman, Ying Hu, and Gianmario Tessitore. Stochastic maximum principle for optimal control of SPDEs. *Appl. Math. Optim.*, 68(2):181–217, 2013.
- [16] D. Grieser. Uniform bounds for eigenfunctions of the Laplacian on manifolds with boundary. *Comm. Partial Differential Equations*, 27(7-8):1283–1299, 2002.
- [17] Ying Hu and Shi Ge Peng. Maximum principle for semilinear stochastic evolution control systems. *Stochastics Stochastics Rep.*, 33(3-4):159–180, 1990.
- [18] Ying Hu and Shi Ge Peng. Adapted solution of a backward semilinear stochastic evolution equation. *Stochastic Anal. Appl.*, 9(4):445–459, 1991.
- [19] Qi Lü and Xu Zhang. General pontryagin-type stochastic maximum principle and backward stochastic evolution equations in infinite dimensions. *preprint arXiv:1204.3275*, 2012.

- [20] Alessandra Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*. Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel, 1995. [2013 reprint of the 1995 original].
- [21] Carlo Orrieri. A stochastic maximum principle with dissipativity conditions. *preprint arXiv:1309.7757*, 2013.
- [22] Carlo Orrieri and Petr Veverka. Necessary stochastic maximum principle for dissipative systems on infinite time horizon. *preprint arXiv*, 2015.
- [23] Shi Ge Peng. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.*, 28(4):966–979, 1990.
- [24] Shan Jian Tang and Xun Jing Li. Maximum principle for optimal control of distributed parameter stochastic systems with random jumps. In *Differential equations, dynamical systems, and control science*, volume 152 of *Lecture Notes in Pure and Appl. Math.*, pages 867–890. Dekker, New York, 1994.
- [25] Gianmario Tessitore. Existence, uniqueness and space regularity of the adapted solutions of a backward SPDE. *Stochastic Anal. Appl.*, 14(4):461–486, 1996.
- [26] J. Yong and X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and emphHJB Equations*. New York: Springer-Verlag, 1999.